UNCLASSIFIED

AD NUMBER
AD006211
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to DoD only; Administrative/Operational Use; DEC 1952. Other requests shall be referred to Personnel Research Laboratory, Lackland AFB, TX. Pre-dates formal DoD distribution statements. Treat as DoD only.**
AUTHORITY
AFHRL ltr dtd 12 Aug 1981

THIS PAGE IS UNCLASSIFIED

THIS REPORT HAS BEEN DECLASSIFIED
AND CLEARED FOR PUBLIC RELEASE.

<u>DISTRIBUTION A</u>
APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

AIR TRAINING COMMAND

HUMAN RESOURCES RESEARCH CENTER

Research Bulletin 52-37

APPLICATION OF THE MULTIPLE DISCRIMINANT FUNCTION TO DATA FROM THE AIRMAN CLASSIFICATION BATTERY

By

DAVID V. TIEDEMAN

JOSEPH G. BRYAN

and

PHILLIP J. RULON

Lackland Air Force Base
San Antonio, Texas
December 1952

NOTIFICATION OF LIMITED DISTRIBUTION

Owing to the limited quantity of color plates available for Research Bulletin 52-37 only 500 copies were printed. This number is considerably less than that required for normal distribution, and it provides no reserve. It has been necessary, accordingly, to reduce the distribution to some agencies in order to maintain adequate distribution to laboratories known to be involved in research for which availability of this report is essential.

APPLICATION OF THE MULTIPLE DISCRIMINANT FUNCTION TO DATA FROM THE AIRMAN CLASSIFICATION BATTERY

Project No. 503-001-0016

Contract No. AF 33(038)-14129

BY

DAVID V. TIEDEMAN
JOSEPH G. BRYAN
and
PHILLIP J. RULON

Educational Research Corporation

3300TH RESEARCH & DEVELOPMENT GROUP
(PERSONNEL RESEARCH LABORATORY)
HUMAN RESOURCES RESEARCH CENTER
AIR TRAINING COMMAND
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

RESEARCH BULLETIN 52-37
DECEMBER 1952

SUBMITTED BY:
LLOYD G. HUMPHREYS
Director of Research
3300th Research & Development Group

IMPLICATIONS FOR THE AIR FORCE

Continuing effort is essential to assure that research on improved selection and classification procedures keeps pace with Air Force requirements. The bulk of past research and development in this area has utilized what is technically termed "correlational analysis." This approach has involved complex statistical analyses of the relationships among measures of aptitudes, abilities, and traits, together with analysis of the relationship of these "predictors" to measures of individual performance in various job activities.

The research described in this report covers development of a new approach, directing attention primarily toward establishment of *differences in abilities* required for various jobs rather than *predictions of success in specific jobs*. This new procedure, utilizing the multiple discriminant function, represents an important supplement to the more traditional correlational analysis. Under certain conditions selection and classification procedures based on multiple discriminant function analysis may yield more definitive outcomes than do procedures based on correlational analysis. Work on application of this technique to Air Force problems is continuing.

This report is highly technical, and will be of interest primarily to research workers. It is of interest to note, however, that development of the multiple discriminant function represents an extension of a more limited technique, developed in Great Britain by R. A. Fisher, and constitutes an important contribution, not only in the field of personnel selection and classification, but in the broader area of mathematical statistics in general. This research bulletin represents the first major publication of this significant work.

Arthur W. Melton

ARTHUR W. MELTON
TECHNICAL DIRECTOR

Herbert M. Cowles

HERBERT N. COWLES
COLONEL, USAF
COMMANDING OFFICER

Headquarters, Air Training Command
Human Resources Research Center
Lackland Air Force Base
San Antonio, Texas

1 December 1952

ACKNOWLEDGMENTS

Dr. John T. Dailey of the Personnel Research Laboratory at the Human Resources Research Center at Lackland Air Force Base participated in the early planning of this study. During conferences conducted by Dr. Dailey decisions were reached concerning the number of groups to be studied, and which groups, and the number of test variables to be included, and which ones. These decisions were of course crucial to the value of the outcomes of the investigation.

Mr. William B. Lecznar of the Personnel Research Laboratory also participated in this planning and, in addition, contributed to various detailed decisions as the work progressed.

Mr. J. Edwin Wade of the Educational Research Corporation was in charge of the computational work, and the nature of the project made this an important responsibility.

Miss Jean Mather, Mr. Raymond Dry, and Mr. Jack Sternberg served as conscientious and facile computers.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF TABLES

APPLICATION OF THE MULTIPLE DISCRIMINANT FUNCTION TO DATA FROM THE AIRMAN CLASSIFICATION BATTERY[1]

INTRODUCTION

A. STATISTICAL TECHNIQUES APPROPRIATE FOR PERSONNEL SELECTION, PLACEMENT, AND GUIDANCE RESEARCH

Certain distinctions among the processes of personnel selection, placement, and guidance have important implications for the type of information required for each. Personnel selection involves the choice of a fixed number of applicants for a job from a pool of applicants which generally exceeds the number chosen. Personnel placement involves the assignment of a fixed pool of employees to several different jobs in which the total number of positions generally equals the number of employees. Personnel guidance involves the consideration of job requirements and personal characteristics within a framework of employment expectations, but the system is not closed in the sense that an available job awaits the client after the guidance interview.

The hiring of a number of applicants for a particular job which is less than the total number who apply as is done in personnel selection requires an estimate of the success of each applicant on the job. After ordering the applicants according to these estimates, those having the highest estimates are hired. Regression analysis is appropriate for defining the success estimates for the particular job. Of course, personnel selection may be complicated by the necessity for hiring, or not hiring, each applicant as interviewed. This complication requires establishment of a minimum estimate of success for the hiring of any applicant. Establishment of the minimum will be influenced by the need for employees and estimates of the rate at which applicants of the desired standard will apply. However, the basic information required is still an estimate of success on the job for which an applicant is interviewed. Regression analysis of previous data will provide such an estimate. The logic of the centour scores, described in the section "Interpretation of Discriminant Scores," is also applicable to this type of problem although in this single-group case one would be unable to compute discriminant functions first.

Personnel placement usually occurs within a framework in which there are the same number of men as positions. The problem consists of achieving the best fit between the capabilities and interests of the men and the requirements of the jobs. Usually this problem is approached as one in which success on each job is estimated for each man and assignments are made by means of these estimates. In order to estimate success, it is necessary to achieve a differentiation of performance within each job and from previous groups to learn how to estimate this differentiation. This knowledge has generally been learned from regression analyses. Some evidence for discontent with this method may be found in the amount of study which has been given to the effect on correlation coefficients obtained in such analyses of previous selection of men. Besides the logical difficulty of expressing success in each job on a common scale as is required for reasoning with estimated success in several jobs, use of the results of regression analyses for personnel placement has the difficulty of continually focusing attention on differentiation within jobs and completely overlooking possible differentiation *among jobs.*[2]

For some years the prevailing purpose of personnel guidance has been to understand both job requirements and individual characteristics and to recommend occupations for which individuals possess the requisite abilities. These principles are also appropriate for personnel placement. We may think of our pool of hired men as possessing different abilities and of the jobs

[1]This report represents a condensation of the complete report submitted by the Educational Research Corporation under Contract AF 33(038)-14129. A limited number of copies of this complete report are available for distribution or loan. [See Bibliography (10).]

[2]This point is amplified in Appendix A of the complete contractor's report on this project (10, pp. 175-200). In the remainder of this condensation the above report will be referred to as the "complete report."

which must be filled as requiring different abilities. Personnel placement involves a matching of human talents and job requirements so as to maximize satisfaction with the total assignment. The satisfaction of both the man and the institution with the assignment may be taken into account in the process.

The result of personnel placement is the attachment of a job label to each man. If a number of characteristics of the men were observed prior to their job assignment, one may study how the personnel placement officer combined this information in making his assignments. For this type of study multiple discriminant analysis is highly appropriate. Multiple discriminant analysis will provide estimates of the number of ways in which the placement officer judged that the jobs with which he was working varied, the manner in which he combined information from his original observations to produce this variation, and the amount of separation of jobs which was achieved by him. Definition of the placement officer's concept of job requirements by multiple discriminant analysis permits accurate communication in case it is desired to reproduce this concept.

Multiple discriminant analysis is a general technique for obtaining group contrasts in terms of a set of n variates. It is not limited to a situation where the groups are defined as arbitrarily as they might be by a placement officer. If further criteria than the placement officer's judgment were introduced into definition of the groups, multiple discriminant analysis would be the most appropriate technique for determining if the groups were differentiated by all n variates. Multiple regression analysis ignores all information the data furnish about the regions of the n-space occupied by the several groups.

Personnel placement and personnel guidance are quite similar processes especially if the philosophy of personnel placement is oriented towards joint satisfaction of the individual and the institution. However, guidance usually operates in a framework in which the availability of specific jobs is either unknown or only crudely estimated. Nevertheless, the same information is needed for personnel guidance as for personnel placement, i.e., information concerning the multivariate distributions of abilities of a large number of occupational groups. Multivariate discriminant analysis will provide an efficient summarization of these multi-dimensional distributions. Multiple regression analysis does not do this because occupations are not contrasted in regression analysis; all the regression analysis achieves is within-group differentiation. Of course, when only successful men in a given occupation are studied, discriminant analysis achieves only among-groups differentiation but it is this type of differentiation which is very much desired in placement and guidance work.

B. CAREER GUIDANCE IN THE AIR FORCE

After World War II the Air Force inaugurated a program of career guidance of basic airmen prior to assignment at the end of indoctrination training. At present, an airman's assignment is based in large part upon his aptitude index scores derived from the Airman Classification Battery.

The Airman Classification Battery AC−1B[3] is based upon the assumptions that there exists a number of human aptitudes and that each job or group of jobs requires a definite patterning of these measured aptitudes. The aptitude indexes of the battery resulted from the research studies done on approximately 100,000 airmen during years 1947 and 1948. For practical use and comparability, the indexes were converted to a normalized nine-point scale (ranging from a low of one to a high of nine points). Each aptitude index is a composite score derived from a combination of weighted tests and Biographical Inventory scores, each of which has proven to be predictive of success in training for, or proficiency in, the Air Force specialties (MOS'S, SSN'S) in one, or more than one, of the eight aptitude clusters or job families.

The eight aptitude clusters corresponding to the job family career fields are;

1. Technician Specialty
2. Mechanical
3. Clerical
4. Equipment Operator
5. Radio Operator
6. Services
7. Crafts
8. Electronics Technician

[3]This (third) form of the Airman Classification Battery became operational 6 Dec 49, following two earlier forms: (1) the preliminary and experimental form of the battery administered prior to 15 Nov 48; (2) operational form (AC-1A) which was in use 15 Nov 48 through 5 Dec 49 (5).

It should be emphasized that the Airman Classification Battery was experimentally tried out and evaluated for two years before it was used operationally in the classification and assignment of airmen. Follow-up studies are made routinely on all airmen assigned to technical schools and the Airman Classification Battery has been, and will continue to be, revised and improved on the basis of these follow-up studies.

Aptitude indexes for each airman are furnished the Indoctrination Wing counselors and assignment officials who are trained, mature men with wide occupational experience and information. The testing record cards are sent to the counselors in advance of the interviews with airmen. In an interview the interests, leisure activities, occupational experience, and educational background of the airman are considerations in making recommendations for assignment. The Air Force Indoctrination Wing is urged to utilize individuals to the capacity of their abilities and interests, within the limits of Air Force needs. There are 309 SSN'S and MOS'S listed specifically under the eight aptitude clusters and airman career fields (5, pp. 229–237). The counselors aid the airmen in deciding upon three career choices. The first two are specific career fields in an aptitude cluster for which an airman has shown qualification and interest. The third is the airman's secondary aptitude cluster and the career field is not specified. Of the basic airmen who attain at least one aptitude index of 5 or higher, 75 per cent to 90 per cent are sent to technical schools after basic training. In 1949 from 5[illegible] per cent to 9[illegible] per cent of the men (for whom data were available) who were sent to technical schools received assignments which were the first recommendation of the career guidance counselors (5, pp. 79). The variations in percentages depend, in a large part, upon priority of the Air Force requirements and the needs which arise for men to fill certain quotas. Also, rigid physical qualifications result in assignment of some airmen otherwise qualified to a second or third recommended career field. Additional research is in progress to determine the effectiveness of the career guidance interviews.

Regardless of whether the Air Force program is a placement or guidance one, the data on which it proceeds have been accumulated mainly by means of regression analysis. Little attention has been given to determining if successful specialists in the many Air Force specialties occupy different regions of the 17-dimensional-Airman-Classification-Battery-space. Since this question is of vital concern for either placement or guidance, the multiple discriminant function technique has been applied to the Airman Classification Battery scores of airmen who have completed study in some one of eight specialties satisfactorily. The technique is described in the complete report (10, pp. 32–128) in detail sufficient to permit independent continuation of this type of research.

TESTS AND AIRMEN STUDIED

Discriminant analysis, like other multivariate analysis techniques, requires a score on each variate for each airman. Since personnel restrictions prior to 15 November 1948 had prevented collection of complete information on many basic airmen, it was decided that only data obtained since this date would be analyzed.

A. THE TESTS

After 15 November 1948, basic airmen were given Airman Classification Battery AC–1 upon induction. The battery is administered and scored by well-trained and well-supervised personnel in a manner described in *Research Bulletin* 50–3 of the Air Training Command's Human Resources Research Center (5, pp. 8–9). Scores on these tests are converted to stanines and aptitude indexes. These scores are then forwarded to the classification and assignment section for use in career counseling interviews.

The following 17 stanine scores were analyzed in this study:

ERC Code	USAF Code	Test Name
	BE601B	Biographical Inventory
1		Clerical Key
2		Mechanical Key
3		Crafts Key
4		Equipment Operator Key
5		Radio Operator Key
6	BI602A	Word Knowledge
7	BI201B	Arithmetic Reasoning
8	BP622A-6[illegible]1A	Dial and Table Reading
9	CI702B	Numerical Operations II
10	BI101B	Aviation Information
11	BI102B	Background for Current Affairs
12	BI901B	Electrical Information
13	BI903B	Mechanical Principles
14	BI902B	General Mechanics
15	BI904B	Tool Functions
16	CP610A	Speed of Identification
17	BI510A	Memory for Landmarks

Sequence of the ERC code was determined by the order in which scores were punched in the IBM cards. Scores for the Instructor key of the Biographical Inventory were not included in the analysis.

Airmen without a score on one or more of these 17 variates were excluded from the analysis.

B. THE AIRMEN

1. Selection of Airmen for Study

After career counseling, airmen who qualified were shipped to a technical training school (5, p. 9).

Choices the airmen expressed during the career counseling interview determined school assignment within restrictions of aptitude indexes, quotas, and personnel supply.

Those airmen who completed one of eight technical schools with an average grade of 2.5 or better were included in the analysis. Airmen who met this criterion were judged to be successful specialists and to constitute a homogeneous group for study.

In this pilot investigation the number of specialties was limited to eight. Selection of the eight specialties by Dr. John T. Dailey, representing the Human Resources Research Center, and Dr. Phillip J. Rulon, representing the Educational Research Corporation, was based on two criteria: (a) the interest, importance, or criticalness of the specialty involved, and (b) the number of cases available for analysis.

2. Specialties Studied

The technical specialties selected for study together with a statement of the number of career guidance cases with technical school results available on 1 July 1950 were as follows:

ERC Code[4]	USAF Course Number	Technical School	Number of Cases Available	Number of Cases Analyzed[5]
A	76601	RO	429	334
B	40500	CT	2124	1966
C	55200	CTO	688	620
D	60000	AMFM	2354	2084[6]
E	75401	RM	667	514
F	78400	WO	371	266
G	55500	ASMW	265	223
H	77501	RaM	107	99
Total	-----	-----------	7005	6106[6]

(Continued)

[In which RO is Radio Operator, CT is Clerk-Typist, CTO is Control Tower Operator, AMFM is Aircraft Maintenance Fundamentals (ATE) Mechanic, RM is Radio Mechanic, WO is Weather Observer, ASMW is Airplane Sheet Metal Worker, and RaM is Radar Mechanic

The importance of the radar mechanic specialty justifies its inclusion despite the small number of cases available.

DISCRIMINATION AMONG EIGHT SPECIALTIES BY MEANS OF THE AIRMAN CLASSIFICATION BATTERY

A. COMPUTATION OF DISCRIMINANT FUNCTIONS

A general description of the application of the multiple discriminant function, including illustrative material, is presented as Appendix A to this report. A less technical description of application of the multiple discriminant function, representing a presentation made by Dr. P. J. Rulon to a research planning conference (8), is included as Appendix B. The present section deals with the application of the multiple discriminant function to data yielded by the Airman Classification Battery.

Computations leading to the derivation of the characteristic equation for the eight-group, seventeen-variate Air Force problem are given in the complete report in Appendix B and in Chapter 4 Tables 4.1 through 4.14 (10).

The characteristic equation is:

$$\lambda^7 - 0.982\,547\,262\,\lambda^6 + 0.186\,697\,242\,\lambda^5 - 0.010\,860\,916\,\lambda^4 + 0.000\,194\,412\,\lambda^3 - 0.000\,001\,288\,\lambda^2 + 0.000\,000\,003\lambda - 0 = 0.$$

The absence of a constant term in the characteristic equation indicates that one of the roots

[4] The sequence was determined largely by the order of receipt of cards in Cambridge, Mass.

[5] All cards were not transmitted, and airmen with missing data or average grade of 2.4 or less were eliminated.

[6] After the analysis was completed the IBM sub-contractor discovered that a duplicate card had been carried through the analysis. This card is included here and in the study reported in the section entitled "Discrimination Among Eight Specialties by Means of the Airman Classification Battery" but excluded in the studies reported in the sections entitled "Discriminant Scores for Each Airman" and "Interpretation of Discriminant Scores."

is $\lambda = 0$. Thus, the discriminant space has no more than six dimensions and the characteristic equation may be reduced to the following sixth-degree equation

$$\lambda^6 - 0.982\,547\,262\,\lambda^5 + 0.186\,697\,242\,\lambda^4 - 0.010\,860\,916\,\lambda^3 + 0.000\,194\,412\,\lambda^2 - 0.000\,001\,288\,\lambda^1 + 0.000\,000\,003 = 0.$$

Roots of this equation were found by synthetic division, a procedure described in Appendix C of the complete report (10).

The trial divisors and correction factors used in the synthetic division process are recorded in Table 4.14 of the complete report (10).[7] The last two roots were obtained by solution of the general quadratic equation. In this solution the radical term proved to be $\sqrt{-0.000\,000\,0[illegible]}$ which was judged to be zero within the accuracy of these computations. The roots were:

Number	Root
1	0.753 434 410
2	0.142 185 849
3	0.062 308 083
4	0.013 618 307
5	0.005 497 623
6	0.005 497 622
7	0

Latent roots define the ratio of among-groups to within-groups sums of squares. Thus they indicate, for each of the possible linear combinations of the original variates, the magnitude of separation of the group centroids relative to a pooled estimate of the dispersion of the individual values about the group means. The root which has been labelled "1," is 5 times as large as Number 2, 12 times as large as Number 3, 58 times as large as Number 4, and 151 times as large as Numbers 5 and 6. In addition root Number 2 is twice as large as Number 3, 11 times as large as Number 4, and 28 times as large as Numbers 5 and 6. Therefore, a considerable amount of information concerning separation of the 17-dimensional centroids for the eight specialties is described by the latent vector associated with root Number 1. Much less information is described by the latent vector associated with root Number 2 and the amount of information described by the latent vectors associated with the remaining five roots is negligible. Reference to Appendix C, Figure 5.9 will make this fact more apparent. The separation of group distributions depicted there is the best which can be achieved by any pair of linear combinations of these 17 variates. Discriminant functions associated with roots 3, 4, 5, and 6 will give even less separation of group centroids and greater overlap of discriminant scores of the airmen.

These considerations when combined with the rapid loss of significant digits which occurred when the roots were raised to the sixth power in the L matrix and when L was premultiplied by B_7, led to abandonment of the last four vectors, leaving only the first two. Thus it was judged that practically all the information concerning separation of the eight 17-dimensional centroids could be described by a two-dimensional space.

[7] Persons interested in a detailed presentation of the multiple discriminant analysis procedure should, of course, refer to the complete report as prepared by the contractor (10), in which Appendix A of this report was presented in Chapter 2.

B. COMPUTATION OF LATENT VECTORS

The computation of the latent vectors associated with each of these two latent roots and the transformation of the vectors back to the original space is presented in Tables 4.15 through 4.18 of the original report (10, pp. 115–119). The unconventionalized discriminant coefficients are given in Table 1 of the present report.

TABLE 1

MATRIX V (UNCONVENTIONALIZED)

	1	2
1.	0.397736914	-0.012659723
2.	-0.287286761	0.003557779
3.	-0.028087056	-0.010373889
4.	-0.036040310	-0.009731979
5.	-0.004583882	0.043487010
6.	0.100139971	0.000400766
7.	0.008725127	0.005163018
8.	0.013220744	0.011312338
9.	0.105039579	0.010214551
10.	-0.109848887	0.007215577
11.	0.009363911	-0.000425362
12.	-0.129596068	0.027894922
13.	-0.034290726	-0.003969322
14.	-0.052145618	-0.004655050
15.	-0.072914445	-0.012631167
16.	-0.009517240	0.001949307
17.	-0.024696097	0.014907050
Sum	-0.154749844	0.071455738
Check	-0.154749845	0.071455738

Each element of a column of Matrix V (unconventionalized) is divided by the largest element in the column. The results of conventionalizing Matrix V in this manner are given in Table 2. Conventionalization changes the units of measurement, of course, but has no effect on

TABLE 2

MATRIX V (CONVENTIONALIZED)

	1	2
1	1.000000000	-0.291115048
2.	-0.722300976	0.081812484
3.	-0.070541745	-0.238551443
4.	-0.080613440	-0.223790484
5.	-0.011524910	1.000000000
6.	0.251774398	0.009215764
7.	0.021936930	0.118728523
8.	0.033239922	0.260131428
9.	0.264093111	0.234887407
10.	-0.276184792	0.165924882
11.	0.023842977	-0.009781358
12.	-0.325833644	0.641634126
13.	-0.086214593	-0.091277586
14.	-0.131108804	-0.107044609
15.	[illegible]	[illegible]
16.	-0.023928480	0.044225041
17.	-0.062091539	0.342793170
Sum	-0.389078891	1.643151324
Check	-0.389075890	1.643151322

separation of groups in terms of the modified units which is the only concern in discriminant analysis.

To test the vectors against their analytic definition, it was necessary to compute AV, WV, and WVΛ to see that AV = WVΛ. Λ is a diagonal matrix with λ_j $(j = 1,2,\ldots,r)$ as the diagonal elements. In the complete report (10), AV, WV, and WVΛ are given in Tables 4.21, 4.22, and 4.23, respectively. Substantial agreement but little identicality was observed in this check. Despite this lack of identicality, the checks were judged satisfactory. The loss of significant digits in discriminant analysis will be discussed in Subsection C, below.

As an additional check, V′AV and V′WV should be computed to test the vanishing of off-diagonal terms. These results are given in Tables 4.24 and 4.25, respectively, in the complete report (10). In both instances the off-diagonal terms were small relative to the diagonal terms.

Finally, the latent roots were tested by taking the quotients of corresponding diagonal elements of V′AV and V′WV. These results were:

Quotient of Corresponding Diagonal Elements	Latent Root
.753 434 401	.753 434 410
.142 193 472	.142 185 848

C. SIGNIFICANT DIGITS IN COMPUTATION

The discriminant analysis method used here, like all methods of numerical analysis, is fraught with problems of significant digits. The causes of these problems are machine capacity, fixed decimals, and limited patience.

The Marchant Automatic Calculating Machine used for these computations was a ten-bank machine with ten-digit quotient capacity. Thus, the maximum accuracy of quotients was limited to ten digits unless quotients were obtained in parts, a procedure which overtaxed the patience of the project staff when it contemplated the number of quotients involved. The process of multiplication on this machine also has a ten-digit capacity for multipliers and multiplicands unless multiplication is done by parts, a procedure which was not used because of the volume of product accumulation required by the method of analysis. Therefore, the maximum number of significant digits considered usable was ten.

Considerable advantage in time, accuracy, and peace of mind in machine computation is achieved by working about a fixed decimal. These considerations led to the use of a fixed decimal so long as serious loss of significant digits did not occur.

The number of decimals carried in the elements of A was determined by the largest of the products of the sums divided by the number of cases. It was found that a maximum of four decimals could be carried. This choice resulted in the elements of A having from six to eight significant digits. Similar considerations led to recording w_{ij} elements to four decimals also. The number of significant digits in W ranged from seven to nine. Since the elements of A and W are involved in multiplication and division processes in the computation of other matrices, it is likely that the number of significant digits in the final Matrix V and the check matrices is not greater than six and may be as low as four or five.

The problem of significant digits is discussed in considerable detail on page 127 of the complete report (10).

It is apparent that some attention must be given to this problem as discriminant analysis progresses. With this attention, however, considerable agreement between original and reproduced numbers will be observed. For these data it is thought that the discriminant function

coefficients have no less than four-digit accuracy and this is as much accuracy as is required for the methods of interpretation which are likely to be used.

DISCRIMINANT SCORES FOR EACH AIRMAN

A. REPORTING RESULTS OF DISCRIMINANT ANALYSIS

The aim of discriminant analysis is to define the smallest space in which information concerning separation of group centroids may be described. The analysis described in the last section indicated that, for all practical purposes, information concerning separation in the Airmen-Classification-Battery space of the eight Air Force specialties studied is contained in two-space. The problem of determining the group-likeness of individual airmen remains, however.

In guidance studies such as this one it is universally found that the various individuals in each group fall at various points in the space of the original variates. This suggests immediately the utility, for guidance, of the concept of probability of a point's being drawn from a particular distribution. However, the use of this concept in the discriminant space requires knowledge of the form of distribution of discriminant scores. Study of this problem was initiated by computing both discriminant scores for each airman.

B. COMPUTATION OF DISCRIMINANT SCORES FOR EACH AIRMAN

Discriminant scores for each airman were computed by IBM machines. Three cards were reproduced for each airman. The cards contained service number, the 17 stanine scores, selected four-digit discriminant function coefficients (taken from Table 2 and multiplied by 1000), and a code indicating whether the accumulated products were positive or negative. Two runs of each deck of the cards through the 602A Multiplying Punch gave partial sums of the products of stanine scores and appropriate discriminant function coefficients. After collation the three partial sums of each discriminant function were summed by the tabulator and punched into a summary card. Since 20000 was added to each discriminant score of each airman to make all scores positive, the resulting discriminant scores were five-digit numbers.[8] Addition of this constant does not affect group separation, of course.

The sum of each of the 17 stanine scores are known for each group. For each discriminant function and for each group the sum of the products of these sums and the corresponding discriminant coefficients must equal the sum of the airmen's discriminant scores. Comparison of the numbers obtained in these two operations provides a check on the machine computation of scores for each airman.

The time required for machine computation of discriminant scores was reduced by using four-digit discriminant function coefficients. It was judged that this accuracy exceeded the accuracy which would be used in interpretation.

C. DESCRIPTION OF DISCRIMINANT SCORES

1. Means, Standard Deviations, and Correlations

Means, standard deviations, and correlations of the two five-digit discriminant function scores were computed for each group. As can be seen in Table 3 the first discriminant function means range from 15.435 to 20.271 and the second from 27.333 to 32.050. Considerable consistency in standard deviations exists from group to group for each discriminant function and the scatter in the second function tends to be larger than it is in the first. Correlations between discriminant functions range from −0.0601 to 0.2554. When testing the hypothesis that correlations were random samples from a population in which $\rho = 0$, the coefficients in groups B (Clerk-Typists), D (Aircraft and Engine Mechanics), E (Radio Mechanics), and H (Radar Mechanics) were found to exceed the .05 level of significance. As is to be expected the correlation of the combined groups is zero within the number of significant digits used in the discriminant coefficients.

2. Bivariate Discriminant Score Distributions

Bivariate distributions of discriminant scores, one for each group, are given in Tables 4 through 11. These tables suggest that the distributions

[8] A detailed statement of the machine procedure may be found in Appendix D of the complete report (10).

TABLE 3

MEANS, STANDARD DEVIATIONS, AND CORRELATION OF DISCRIMINANT SCORES BY GROUP

Group	No. of Cases	Mean 1	Mean 2	SD 1	SD 2	Correlation (r)	$\frac{r}{\sigma_r}$
A	334	18.198	31.455	2.277	2.380	-.0042	-0.073
B	1966	20.221	29.238	2.344	2.624	-.0480	-2.124*
C	620	18.142	31.202	2.497	2.433	-.0132	-0.323
D	2083	15.435	29.298	2.416	2.579	.0544	2.466*
E	514	16.041	30.818	2.217	2.529	.1213	2.738*
F	266	18.710	30.367	2.538	2.933	-.0681	-1.107
G	223	16.163	27.333	2.030	2.462	.0144	0.209
H	99	16.079	32.050	2.390	2.066	.2554	2.525*
Combined	6105	17.633	29.737	3.138	8.685	-.00008	--

*Exceeds .05 level of 1.96 at $\rho = o$. In each instance σ_r was computed under the hypothesis, $\rho = o$.

TABLE 4

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP A (RADIO OPERATORS)

Discriminant Function 1

Class Index (Discriminant Function 2)	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	–	1	–	–	–	–	1	2	–	1	–	–	–	–	–	–	5
36	–	–	–	–	1	–	1	1	1	4	1	2	1	–	–	–	–	–	–	–	12
35	–	–	–	–	–	3	–	2	6	3	4	1	1	–	2	–	–	–	–	–	22
34	–	–	–	–	1	–	4	5	4	3	4	7	2	1	1	–	–	–	–	–	32
33	–	–	–	–	1	2	5	5	3	3	4	2	5	3	2	–	–	–	–	–	35
32	–	–	–	–	2	3	3	5	6	13	6	8	3	3	2	–	–	–	–	–	54
31	–	–	–	–	–	–	5	7	4	8	14	10	6	2	1	–	–	–	–	–	57
30	–	–	–	–	2	1	5	4	6	9	12	9	2	1	–	–	–	–	–	–	51
29	–	–	–	–	1	1	1	3	2	3	3	4	3	3	1	1	–	–	–	–	26
28	–	–	–	–	1	–	2	–	2	7	5	2	2	–	–	–	–	–	–	–	21
27	–	–	–	–	–	1	–	1	4	4	3	–	2	–	–	–	–	–	–	–	15
26	–	–	–	–	–	–	–	1	–	1	1	–	1	–	–	–	–	–	–	–	4
25	–	–	–	–	–	–	–	–	–			–	–	–	–	–	–	–	–	–	–
24	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
23	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	–	–	–	9	12	26	34	38	58	58	47	28	14	9	1	–	–	–	–	334

are roughly bivariate normal. Also since the tables have a common scale, some idea of the overlap in the distributions can be obtained by rapidly leafing through all eight tables. The overlap is demonstrated more clearly in Figure C-1 in Appendix C.

3. *Study of Normality*

The fact that the original scores on which these discriminant functions are based would tend to be approximately normally distributed in large unselected samples, and the fact that discriminant scores are the sums of 17 of these elements, suggested that the central limit theorem might operate in these data to produce bivariate normal distributions in each group. Consequently, tests were made to ascertain whether it happened or not.

A necessary and sufficient condition of bivariate normality is that each of the marginal distributions be normal. Chi square was computed for each marginal distribution to determine if each might be considered a random sample from a normal distribution.

Results of these tests are given in Table 12. Only the Chi squares for the first discriminant scores of Group D (Aircraft and Engine Mechanics) and the second discriminant scores of Groups B (Clerk-Typists) and D (Aircraft and Engine Mechanics) were so large that they would appear less than 5 times in 100 in random samples from a normal distribution.

Since the Chi square test is relatively insensitive to small variations in skewness and kurtosis of a distribution, Fisher's g-statistics of skewness and kurtosis (4, p.75) were also computed for these data. Results of these computations are given in Table 13. The skewness of the first discriminant function departs significantly from the skewness of a normal distribution in Group D (Aircraft and Engine Mechanics) and that of the second in Groups D (Aircraft and Engine Mechanics) and G (Airplane Sheet Metal Workers). Significant departures from the

TABLE 5

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP B (CLERK-TYPISTS)

Discriminant Function 2 Class Index	Discriminant Function 1 9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	–	–	–	–	1	–	1	–	–	1	–	–	–	–	–	–	3
36	–	–	–	–	–	–	–	–	4	2	5	3	1	1	1	1	1	–	–	–	19
35	–	–	–	–	–	–	–	–	2	6	6	4	8	5	1	–	–	1	–	–	30
34	–	–	–	–	–	2	1	5	6	5	7	11	7	8	5	5	1	–	–	–	60
33	–	–	–	–	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	19	14	23	19	12	8	7	2	1	–	–	120
32	–	–	–	1	1	2	3	7	10	16	17	36	27	12	18	7	4	–	–	–	161
31	–	–	–	1	1	3	4	11	12	25	23	38	32	33	17	16	7	1	–	–	224
30	–	–	–	–	–	–	4	5	15	31	32	49	42	29	24	15	7	4	–	–	256
29	–	–	–	–	1	1	7	9	16	20	50	50	49	47	20	12	9	1	–	–	292
28	–	–	–	–	–	–	–	7	13	29	35	45	33	36	25	22	11	2	–	1	259
27	–	–	–	–	–	1	1	9	16	36	45	29	42	34	18	12	1	2	1	–	247
26	–	–	–	–	–	1	–	6	12	13	33	28	20	16	15	4	3	2	1	–	154
25	–	–	–	–	–	1	–	1	3	9	18	13	11	17	10	2	1	–	–	–	86
24	–	–	–	–	–	–	–	1	–	6	10	12	2	8	4	1	–	–	–	–	44
23	–	–	–	–	–	–	–	–	–	1	2	2	1	–	1	–	–	–	–	–	7
22	–	–	–	–	–	–	–	1	–	–	–	–	–	–	1	–	1	–	–	–	3
21	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	1
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	–	–	2	4	15	24	68	118	211	298	343	291	255	168	104	48	14	2	1	1966

TABLE 6

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP C (CONTROL TOWER OPERATORS)

Discriminant Function 2 Class Index	Discriminant Function 1 9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
39	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	–	–	–	–	4	1	–	–	–	–	–	–	–	–	–	–	5
36	–	–	–	–	–	–	1	3	2	6	2	2	3	–	–	–	–	–	–	–	19
35	–	–	–	–	–	2	6	3	–	5	8	2	–	2	–	2	–	–	–	–	27
34	–	–	–	–	2	5	3	11	11	14	6	4	6	1	2	–	–	–	–	–	65
33	–	–	1	1	3	3	7	5	15	10	14	10	6	2	2	1	1	–	–	–	81
32	–	–	–	1	1	4	5	18	11	16	16	2	1	1	3	1	–	–	–	–	80
31	–	–	1	1	2	4	5	10	11	13	15	12	7	10	7	2	–	1	–	–	101
30	–	–	–	–	–	1	5	13	15	15	9	9	9	5	1	1	1	–	–	–	84
29	–	–	–	1	–	4	5	6	7	11	9	10	4	6	2	1	–	–	–	–	66
28	–	–	–	2	–	1	3	6	4	14	10	5	3	5	1	–	–	–	–	–	54
27	–	–	–	–	–	–	1	2	2	7	4	1	3	–	3	–	–	–	–	–	23
26	–	–	–	–	–	1	–	–	1	2	2	2	–	1	–	–	–	–	–	–	9
25	–	–	–	–	–	–	–	–	1	1	1	–	–	–	–	–	–	–	–	–	3
24	–	–	–	–	–	–	–	–	–	1	–	–	–	–	1	–	–	–	–	–	2
23	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	–	3	6	8	25	41	77	84	116	93	59	42	33	22	8	2	1	–	–	620

TABLE 7

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP D (AIRCRAFT AND ENGINE MECHANICS)

Discriminant Function 2 Class Index	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	1
37	–	–	–	–	1	–	1	–	–	–	1	1	–	–	–	–	–	–	–	–	4
36	–	–	–	–	–	–	2	3	5	2	1	1	–	–	–	–	–	–	–	–	14
35	–	–	–	2	2	5	3	3	7*	4	3	1	2	–	–	–	–	–	–	–	32
34	–	–	1	3	10	12	9	8	9	8	6	1	2	2	–	–	–	–	–	–	71
33	–	–	6	6	12	20	17	19	16	6	7	4	2	1	1	–	–	–	–	–	117
32	–	–	5	13	17	25	28	24	29	15	7	3	2	1	1	–	1	1	–	–	172
31	1	2	6	20	32	46	48	37	27	21	18	4	3	1	1	1	–	–	–	–	268
30	1	3	9	26	31	46	47	39	30	25	15	9	3	2	–	–	–	–	–	–	286
29	–	3	18	21	29	41	53	39	33	24	11	10	6	3	1	2	1	–	–	–	295
28	–	10	6	15	22	40	50	41	31	29	13	9	3	1	2	–	–	–	–	–	272
27	–	1	5	18	21	41	45	35	38	17	19	7	1	1	1	–	–	–	–	–	250
26	1	1	5	13	10	26	22	33	23	13	2	2	2	3	–	–	–	–	–	–	157
25	–	1	–	2	12	15	16	18	13	6	2	1	2	1	–	–	–	–	–	–	89
24	–	–	–	3	6	10	15	6	3	1	1	–	–	–	–	–	–	–	–	–	45
23	–	–	–	–	–	1	3	1	1	1	–	–	1	–	–	–	–	–	–	–	8
22	–	–	–	–	–	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	1	21	61	142	205	330	360	306	265	172	106	54	29	16	7	3	2	1	–	–	2083

Columns: Discriminant Function 1

TABLE 8

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP E (RADIO MECHANICS)

Discriminant Function 2 Class Index	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	2	–	1	2	–	–	1	1	1	–	–	–	–	–	–	–	8
36	–	–	–	–	–	–	1	1	1	4	–	–	–	1	–	–	–	–	–	–	8
35	–	–	–	–	2	2	3	5	2	1	1	1	1	1	–	–	–	–	–	–	19
34	–	–	–	1	–	1	5	8	7	6	–	2	2	1	–	–	–	–	–	–	33
33	–	–	–	2	6	9	9	12	10	8	6	2	2	1	1	–	–	–	–	–	68
32	–	–	1	4	4	12	16	11	8	5	9	2	–	1	–	–	–	–	–	–	73
31	–	–	3	2	2	16	15	15	13	7	5	2	1	–	–	–	–	–	–	–	81
30	–	1	1	3	6	7	14	11	12	12	11	–	–	–	–	–	–	–	–	–	76
29	–	1	1	3	2	10	8	7	8	5	4	1	–	2	–	–	–	–	–	–	52
28	–	–	1	1	3	7	3	5	9	7	3	1	–	–	–	–	–	–	–	–	40
27	–	–	–	1	3	1	5	3	5	4	2	–	–	–	–	–	–	–	–	–	24
26	–	–	–	–	3	1	5	3	2	2	2	–	–	–	–	–	–	–	–	–	18
25	–	–	–	2	–	2	2	1	–	3	–	–	–	–	–	–	–	–	–	–	10
24	–	–	–	–	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	2
23	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	2	7	19	33	60	88	85	77	64	44	12	7	7	1	–	–	–	–	–	514

Columns: Discriminant Function 1

kurtosis of a normal distribution were found on the first discriminant function in Group D (Aircraft and Engine Mechanics) and on the second in Groups B (Clerk-Typists), D (Airplane and Engine Mechanics), and F (Weather Observers). Thus more information concerning the non-normality of first discriminant scores in Group B (Clerk-Typists) and D (Aircraft and Engine Mechanics) has been obtained and, in addition, it is apparent that the second discriminant function is non-normal in Groups F (Weather Observers) and G (Airplane Sheet Metal Workers). The extent of the departures from normality may be observed in Figures 5.1 through 5.9 presented in Appendix C in which normal distributions with means and standard deviations of those of the sample are superimposed on the sample histograms. It would seem that the assumption of bivariate normality is not too valid for Group D (Aircraft and Engine Mechanics) but that it is not too inaccurate a representation of the actual situation in the other groups.

4. *Overlap in Discriminant Score Distributions*

Although the discriminant scores of Group D (Aircraft and Engine Mechanics) do not satisfy

the condition of bivariate normality well, this assumption was used in describing overlap in the group distributions. An assumption of this nature was necessary to simplify visualization of the separation of groups effected by the discriminant functions. Recording the actual frequency of occurrence of pairs of scores in each group on a single plane became so confusing that the essential information of the table was lost.

In a bivariate normal population, Rietz (9, p.106) gives the probability that $(y_1-\bar{y}_1)$ will fall in dy_1 and the corresponding $(y_2-\bar{y}_2)$ in dy_2 as

$$\frac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}e^{-A}\,dy_1dy_2$$

where A =

$$\frac{1}{2(1-r^2)}\left[\frac{(y_1-\bar{y}_1)^2}{\sigma_1^2}+\frac{(y_2-\bar{y}_2)^2}{\sigma_2^2}-2r\frac{(y_1-\bar{y}_1)(y_2-\bar{y}_2)}{\sigma_1\sigma_2}\right]$$

The maximum ordinate is

$$z_m = \frac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}$$

TABLE 9

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP F (WEATHER OBSERVERS)

Discriminant Function 2 (rows); Discriminant Function 1 (columns)

Class Index	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	1
37	–	–	–	–	–	–	–	1	–	–	–	1	–	–	–	–	–	–	–	–	2
36	–	–	–	–	–	–	–	–	–	3	2	–	1	–	–	–	–	–	–	–	6
35	–	–	–	–	–	–	–	3	2	–	1	–	1	2	–	1	–	–	–	–	10
34	–	–	–	–	–	2	1	3	5	3	2	1	4	–	–	–	1	–	–	–	22
33	–	–	–	–	–	4	3	2	2	2	7	6	1	3	–	–	–	–	–	–	30
32	–	–	–	–	1	2	1	2	5	3	5	4	5	1	2	–	–	–	–	–	31
31	–	–	–	–	–	2	3	2	5	3	5	3	1	2	2	1	–	–	–	–	29
30	–	–	–	–	1	1	–	3	4	3	5	3	4	3	2	2	–	1	–	–	32
29	–	–	–	–	1	–	–	2	6	5	4	5	1	4	1	1	–	–	–	–	30
28	–	–	–	–	2	–	2	1	1	–	4	4	6	1	2	–	–	–	–	–	23
27	–	–	–	–	–	1	1	1	3	5	3	4	3	1	–	–	–	–	–	–	22
26	–	–	–	–	–	–	–	2	1	2	4	2	2	1	–	–	–	–	–	–	14
25	–	–	–	–	–	–	1	–	1	–	1	2	1	–	1	–	–	–	–	–	7
24	–	–	–	–	–	–	–	1	–	3	1	1	–	1	–	–	–	–	–	–	7
23	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	–	–	–	5	12	12	23	35	32	44	36	31	19	10	5	1	1	–	–	266

TABLE 10

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP G (AIRPLANE SHEET METAL WORKERS)

Discriminant Function 2 (rows); Discriminant Function 1 (columns)

Class Index	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
36	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
35	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
34	–	–	–	–	–	–	–	1	1	1	–	–	–	–	–	–	–	–	–	–	3
33	–	–	–	–	–	1	–	–	3	2	1	2	–	–	–	–	–	–	–	–	9
32	–	–	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2
31	–	–	–	–	1	1	1	–	3	–	1	–	–	–	–	–	–	–	–	–	7
30	–	1	–	1	–	1	3	7	1	3	2	1	–	–	–	–	–	–	–	–	20
29	–	–	1	–	3	2	4	2	6	4	3	–	1	–	–	–	–	–	–	–	26
28	–	–	–	–	2	2	6	6	8	6	6	–	1	–	–	–	–	–	–	–	37
27	–	–	–	–	4	3	3	11	4	7	–	1	–	–	–	–	–	–	–	–	33
26	–	–	1	1	5	4	6	8	2	1	2	–	–	–	–	–	–	–	–	–	30
25	–	–	–	2	1	5	2	5	4	9	1	–	2	–	–	–	–	–	–	–	31
24	–	–	–	–	–	1	4	5	2	1	–	–	–	–	–	–	–	–	–	–	13
23	–	–	1	–	–	–	2	5	2	1	–	–	–	–	–	–	–	–	–	–	11
22	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	1
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	1	3	4	17	21	31	51	36	35	16	4	4	–	–	–	–	–	–	–	223

TABLE 11

BIVARIATE DISCRIMINANT SCORE DISTRIBUTION GROUP H (RADAR MECHANICS)

Discriminant Function 2 — Class Index	Discriminant Function 1: 9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	–	–	–	2	–	–	–	–	–	–	–	–	–	–	–	–	2
36	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	1
35	–	–	–	–	1	–	2	2	2	1	2	–	2	–	–	–	–	–	–	–	12
34	–	–	1	–	–	1	3	1	3	1	1	–	1	–	–	–	–	–	–	–	12
33	–	–	1	–	1	2	1	2	1	2	–	1	1	–	–	–	–	–	–	–	12
32	–	–	–	1	3	1	3	5	3	1	–	3	–	–	–	–	–	–	–	–	20
31	–	–	1	–	–	1	1	6	4	1	–	1	–	–	–	–	–	–	–	–	15
30	–	–	–	1	–	5	2	1	1	2	1	1	–	–	–	–	–	–	–	–	14
29	–	–	–	–	1	–	1	2	1	1	–	–	–	–	–	–	–	–	–	–	6
28	–	–	1	–	–	–	1	1	1	–	–	–	–	–	–	–	–	–	–	–	4
27	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
26	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
25	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
24	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
23	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Total	–	–	5	2	6	10	14	22	16	9	4	7	4	–	–	–	–	–	–	–	99

TABLE 12

CHI SQUARE TEST OF NORMALITY OF DISCRIMINANT SCORE DISTRIBUTIONS FOR EACH GROUP

Group	Function	Number	Chi Square	Number of Degrees of Freedom*	p
A	1	334	10.4299	9	.4003
	2		9.1375	9	.4227
B	1	1966	12.5020	13	.5283
	2		32.1022	14	.0057
C	1	620	15.2735	13	.2946
	2		12.1824	11	.3564
D	1	2083	56.5290	12	.0000
	2		25.0233	14	.0388
E	1	514	13.3523	11	.2735
	2		10.73[illegible]	11	.4665
F	1	266	5.6349	10	.8404
	2		6.5497	11	.8293
G	1	223	6.5578	6	.3675
	2		12.1853	8	.1478
H	1	99	9.6443	8	.2913
	2		7.5953	8	.4762

*The marginal frequencies reported in Tables 4 through 11 were used for these computations.

which occurs when $(y_1-\bar{y}_1) = (y_2-\bar{y}_2) = 0$. Ordinates which are k times z_m, $0<k<1$, lie on the ellipse defined by

$$\sigma_2^2(y_1-\bar{y}_1)^2 + \sigma_1^2(y_2-\bar{y}_2)^2 - 2r\sigma_1\sigma_2(y_1-\bar{y}_1)(y_2-\bar{y}_2) + 2(1-r^2)\sigma_1^2\sigma_2^2 \ln k = 0 .$$

The center of this ellipse is $\bar{y}_1, \bar{y}_2$.

When rotated through an angle, φ, defined by

$$\tan 2\varphi = \frac{-2r\sigma_1\sigma_2}{\sigma_2^2-\sigma_1^2} , \qquad (1)$$

the semi-axis of the ellipse in the direction to which y_1 is rotated is given by

$$\left[\frac{-4(1-r^2)\,Q\,\sigma_1^2\sigma_2^2 \ln k}{(\sigma_2^2-\sigma_1^2)^2 + Q(\sigma_2^2+\sigma_1^2) + 4r^2\sigma_1^2\sigma_2^2}\right]^{1/2} \qquad (2)$$

where

$$Q = \sqrt{(\sigma_2^2-\sigma_1^2)^2 + 4r^2\sigma_1^2\sigma_2^2} \qquad (3)$$

TABLE 13

SKEWNESS AND KURTOSIS OF DISCRIMINANT SCORE DISTRIBUTIONS FOR EACH GROUP

Group	Function	Number	Mean*	Sigma	g_1	σ_{g_1}	$\frac{g_1}{\sigma_{g_1}}$	g_2	σ_{g_2}	$\frac{g_2}{\sigma_{g_2}}$
A	1	334	12.1976	2.3182	-.1312	.1340	- .9799	-.4100	.2680	-1.5300
	2		31.4552	2.4031	.0529		.3948	-.4808		-1.7940
B	1	1966	20.2375	2.3652	-.1022	.0552	-1.8514	.1151	.1104	[illegible]
	2		29.2309	2.6389	.0166		.3007	-.2367		-2.1440§
C	1	620	19.1468	2.5227	.1035	.0984	[illegible]	[illegible]	.1968	.0356
	2		31.2113	2.4578	-.0204		- .2073	-.3381		-1.7180
D	1	2083	15.4583	2.4435	.3928	.0537	7.3147†	.3787	.1074	3.5261§
	2		29.2909	2.6084	.1383		2.5754†	.2389		2.2344§
E	1	514	16.0447	2.2508	.1468	.1080	1.3593	-.0094	.2160	- .0435
	2		30.8249	2.5553	.1157		1.0713	-.1580		- .7315
F	1	266	18.7143	2.5505	-.0378	.1502	- .2517	-.3517	.3004	-1.1708
	2		30.3835	2.9627	-.0754		- .5020	-.6456		-2.1491§
G	1	223	16.1300	2.0347	[illegible]	.1640	[illegible]	[illegible]	.3280	.1174
	2		27.3677	2.5019	.4126		2.5159†	-.0301		- .0092
H	1	99	16.0909	2.3873	-.0146	.2462	- .0593	-.1833	.4924	- .3723
	2		32.0404	2.1269	.0214		.0869	-.4857		- .9864

*The marginal frequencies reported in Tables 4 through 11 were used for these computations.

†Exceeded .05 level of 1.96 at $\gamma_1 = 0$.

§Exceeded .05 level of 1.96 at $\gamma_2 = 0$.

TABLE 14

ANGLE OF ROTATION AND SEMI-AXES OF GAUSSIAN CONTOUR ELLIPSES AFTER ROTATION OF EACH PAIR OF AXES

Group	Angle of Rotation	k = 0.9		k = 0.7		k = 0.5		k = 0.3		k = 0.1	
		1*	2†	1*	2†	1*	2†	1*	2†	1*	2†
A	0°16′	1.05	1.09	1.92	2.01	2.68	2.80	3.53	3.69	4.89	5.11
B	11°30′	1.07	1.21	1.97	2.23	2.74	3.10	3.62	4.09	5.00	5.66
C	13°30′	1.12	1.15	2.05	2.11	2.86	2.94	3.77	3.88	5.21	5.37
D	10°50′	1.10	1.20	2.02	2.20	2.81	3.06	3.71	4.04	[illegible]	[illegible]
E	21°20′	0.99	1.18	1.87	2.18	2.54	3.04	3.35	4.00	4.63	5.54
F	12°30′	1.16	1.36	2.12	2.49	2.96	3.48	3.90	4.58	5.40	6.34
G	2°6′	0.93	1.13	1.71	2.08	2.39	2.90	3.15	3.82	4.36	5.28
H	30°10′	1.17	0.86	2.14	1.59	2.99	2.22	3.94	2.92	5.45	4.04

*Semi-axis of Discriminant Function 1 after rotation to new axes.

†Semi-axis of Discriminant Function 2 after rotation to new axes.

The semi-axis of the ellipse in the direction to which y_2 is rotated is given by

$$\left[\frac{-4(1-r^2)\, Q\, \sigma_1^2\, \sigma_2^2\, \ln k}{-(\sigma_2^2-\sigma_1^2)^2 + Q(\sigma_2^2+\sigma_1^2) - 4r^2\, \sigma_1^2\, \sigma_2^2}\right]^{1/2} \quad (4)$$

Density distributions may be depicted in a plane by drawing contour lines representing a constant increment in the height of the probability surface. Accordingly, five contour ellipses were chosen for each group such that the difference in height of the probability surface was .2 between each contour and the next. These ellipses were defined by the above relationships by choosing values of k at 0.9, 0.7, 0.5, 0.3, and 0.1.

The means, standard deviations, and correlations reported in Table 3 were used in computation of the angle of rotation by equation (1) and the semi-axes by equations (2), (3), and (4). The results of these computations are reported in Table 14.

The Gaussian contour ellipses are shown in Figure 5.9 (Appendix C). This figure demonstrates the overlap in abilities represented in airmen who have successfully completed training in one of the eight specialties which were studied. It must also be remembered that the differentiation demonstrated in this figure is achieved by the two discriminant functions with the highest latent roots. If similar contour ellipses were constructed for the discriminant functions associated with any pair of the remaining four non-zero latent roots or for any combination of these four with one of the two shown in Figure 5.9 (Appendix C), greater overlap would be observed in the distributions.

Contours for individual groups may be distinguished in Figure 5.9 (Appendix C) if a color fixation is consciously assumed while looking at the figure. The distinction is aided if the center of the contours is first identified. Contour centers are identified in Figure 5.9 (Appendix C) by the centers of the large crosses (+). Contour centers occur in a region bounded by Discriminant Function 1 scores from 15 to 20 and by Discriminant Function 2 scores from 27 to 33. For instance, the center of the solid-green (radar mechanics) contours is at approximately a Discriminant Function 1 score of 16 and a Discriminant Function 2 score of 32. If this cross is located the five solid-green ellipses surrounding it can be identified with very little effort if a green-consciousness is assumed by the reader. Some of the differentiation of the groups which actually exists in the data can be seen if the solid-green (radar mechanics) contours are contrasted with both the broken-green (clerk-typists) contours and the broken-red (sheet metal workers) contours. Identity of contours is closest in the solid-red (radio operators) contours and the broken-blue (control tower operators) contours.

Further information may be obtained from Figure 5.9 (Appendix C) if one identifies the contour ellipses on which various pairs of discriminant scores lie. For instance, a Discriminant Function 1 score of approximately 12 and a Discriminant Function 2 score of approximately 28 lies on the solid-blue (A & E mechanics) ellipse with k=0.3, the broken-red (sheet metal workers) ellipse with k=0.1, and the broken-orange (weather observers) ellipse with k=0.1. This point is outside even the k=0.1 ellipses of the other five groups. Thus the relative frequency with which this score occurred was higher in the group of airmen assigned to A & E mechanic school than it was in any of the other seven groups.

INTERPRETATION OF DISCRIMINANT SCORES

A. PROBABILITY STATEMENTS CONCERNING GROUP MEMBERSHIP

Suppose that the density function of discriminant scores were known for a particular population. With this information it is possible to compute the probability that a random point, $y_1, y_2, \ldots, y_r$ let us say, will be in the small interval $dy_1, dy_2, \ldots, dy_r$. It is also possible to determine the density outside a given boundary. This latter information would specify the fraction of the total density beyond the particular boundary on which $y_1, y_2, \ldots, y_r$ lies.

Suppose further, that a contour of the density were chosen such that the fraction α of the population of Group A individuals is outside the contour. If a random sample of Group A individuals were classified as A or not-A on the basis of whether their scores placed them inside or outside this contour, the expectation is that the fraction α of individuals who were really Group A individuals would not be called Group A individuals. The fraction α associated with a

given centour is thus the probability of misclassifying an individual as "not-A" when he really belongs to Group A. This fraction, when computed from a sample and expressed as a per cent, will be called a *centour score*.[9] A centour score specifies the risks involved in acting as if a person is not a member of a particular group when in reality he is. The importance of centour scores for career guidance is obvious.

In a previous section it was reported that the discriminant scores were approximately bivariate normal in all groups with the exception of Group D (Aircraft and Engine Mechanics). Despite this exception, centour scores for this group as well as for the other groups were computed as though the density function was bivariate normal. It was judged that this assumption would yield in each group the one- or possibly two-digit accuracy necessary for interpretation of the scores in guidance conferences. Sample means, standard deviations, and correlations reported in Table 3 were used in computing centour scores, since they are maximum likelihood estimates of their respective population parameters in a sample from a bivariate normal population.

Rietz states:

"The equal-frequency curves obtained by making z take constant values in

$$z = \frac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}\, e^{-B},$$

where B =

$$\frac{1}{2(1-r^2)}\left[\frac{(y_1-\bar{y}_1)^2}{\sigma_1^2}+\frac{(y_2-\bar{y}_2)^2}{\sigma_2^2}-2r\frac{(y_1-\bar{y}_1)(y_2-\bar{y}_2)}{\sigma_1\sigma_2}\right]$$

are an infinite system of homothetic ellipses, any one of which has an equation of the form

$$\frac{(y_1-\bar{y}_1)^2}{\sigma_1^2}+\frac{(y_2-\bar{y}_2)^2}{\sigma_2^2}-2r\frac{(y_1-\bar{y}_1)(y_2-\bar{y}_2)}{\sigma_1\sigma_2}=\lambda^2.$$

Further,..... "The probability that a point (y_1, y_2) taken at random will fall within any ellipse obtained by assigning λ is given by

$$1 - e^{-\frac{\lambda^2}{2(1-r^2)}}$$ " (9, p.108)

Conversely, the probability, p, that a point will fall outside this ellipse is

$$p = e^{\frac{-\lambda^2}{2(1-r^2)}}$$

From this we see that

$$\lambda^2 = -2(1-r^2)\ln p$$

and the ellipse outside of which the proportion p of the area will fall is

$$\frac{(y_1-\bar{y}_1)^2}{\sigma_1^2}+\frac{(y_2-\bar{y}_2)^2}{\sigma_2^2}-2r\frac{(y_1-\bar{y}_1)(y_2-\bar{y}_2)}{\sigma_1\sigma_2} = -2(1-r^2)\ln p.$$

Thus

$$p = \text{Antiln}\left\{-C\right\}, \qquad (5)$$

where $-C$ =

$$\frac{1}{2(1-r^2)}\left[\frac{(y_1-\bar{y}_1)^2}{\sigma_1^2}+\frac{(y_2-\bar{y}_2)^2}{\sigma_2^2}-2r\frac{(y_1-\bar{y}_1)(y_2-\bar{y}_2)}{\sigma_1\sigma_2}\right]$$

and the centour score associated with y_1, y_2 is 100p. Since the multivariate normal density function has been generalized to n variates, the centour score concept may also be generalized to n discriminants. Such a generalization has been accomplished by Pearson (7, pp.xxiv-xxvi).

B. CENTOUR SCORES FOR GUIDANCE OF AIRMEN

Centour scores were computed for each pair of discriminant scores for each group.[10] They are reported in Table 15.

Suppose that an airman is given the Airman Classification Battery, his tests are scored, and his two discriminant scores are computed.[11] Let us say that Discriminant Function 1 score proves to be 19 and Discriminant Function 2 score proves to be 24. From Table 15 we see that centour scores for this pair of discriminant scores are 1,11,1,3,1,9,15, and * for Groups A, B,C,D,E,F,G, and H, respectively. The guidance counselor could now say to the airman:

1. If I said that airmen like you do not satisfactorily complete Radio Operator (Group A) school, I would be excluding 1 per cent of the group who did so.
2. If I said that airmen like you do not satisfactorily complete Clerk-Typist (Group B) school, I would be excluding 11 per cent of the group who did so.

[9] The word, "centour" was coined by combining the "cent" of "per cent" with the "our" of contour. It was thought that "centour" adequately conveyed the idea, "per cent of area beyond the contour on which a point lies." Since centour scores have been computed by assuming a multivariate normal distribution in each group, the Air Force could continue the stanine frame of reference it has chosen to maintain by reporting "stanours" instead of centours. The same areas used to define stanines could be used to define "stanours."

[10] See Appendix D of the complete report (10) for computational routine.

[11] The coefficients of the discriminant scores for these computations may be found in Table 2. Twenty (20) must be added to each of these scores before entering Table 15.

3. If I said that airmen like you do not satisfactorily complete Control Tower Operator (Group C) school, I would be excluding 1 per cent of the group who did so.
4. If I said that airmen like you do not satisfactorily complete Aircraft and Engine Mechanic (Group D) school, I would be excluding 3 per cent of the group who did so.
5. If I said that airmen like you do not satisfactorily complete Radio Mechanic (Group E) school, I would be excluding 1 per cent of the group who did so.
6. If I said that airmen like you do not satisfactorily complete Weather Observer (Group F) school, I would be excluding 9 per cent of the group who did so.
7. If I said that airmen like you do not satisfactorily complete Airplane Sheet Metal Worker (Group G) school, I would be excluding 15 per cent of the group who did so.
8. If I said that airmen like you do not satisfactorily complete Radar Mechanic (Group H) school, I would be excluding less than 0.5 per cent of the group who did so.

If I don't mind being wrong 5 per cent of the time, I would say that you're *not* like the airmen who satisfactorily completed radio operator, control tower operator, aircraft and engine mechanic, radio mechanic, and radar mechanic schools. You look more like other airmen who have satisfactorily completed either clerk-typist, weather observer, or airplane sheet metal worker schools. If I had to choose one of these schools for you on the basis of these 17 tests, I would select airplane sheet metal worker school, since the probability is largest for your belonging to this group.

In Figure 6.1 (Appendix C) the regions where centour scores of each group are maximal are mapped. This figure demonstrates quite clearly the consequences for specialty classification of overlapping distributions. If assignment to specialties is done entirely in terms of the

(Text continues on page 20)

TABLE 15

CENTOUR SCORES EIGHT AIR FORCE SPECIALTIES

Discriminant Function 1(+20)

Discriminant Function 2(+20)	Group†	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	B	•	•	•	•	•	•	•	•	•	•	1	1	1	1	•	•	•	•	•	•
	C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
21	D	•	•	•	•	•	•	1	1	•	•	•	•	•	•	•	•	•	•	•	•
	E	•	•	•	•	•	•	•	•	•	•	1	1	•	•	•	•	•	•	•	•
	F	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	G	•	•	•	•	1	2	3	4	3	2	1	1	•	•	•	•	•	•	•	•
	H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	B	•	•	•	•	•	•	•	•	1	1	2	2	2	2	1	1	•	•	•	•
	C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
22	D	•	•	•	1	1	2	2	2	1	1	•	•	•	•	•	•	•	•	•	•
	E	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	F	•	•	•	•	•	•	•	1	1	2	2	2	1	1	1	•	•	•	•	•
	G	•	•	•	1	3	6	8	10	9	6	3	2	1	•	•	•	•	•	•	•
	H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	B	•	•	•	•	•	•	•	1	2	3	3	6	6	5	3	2	1	•	•	•
	C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
23	D	•	1	1	2	3	5	5	5	4	2	1	1	•	•	•	•	•	•	•	•
	E	•	•	•	•	1	1	1	1	1	•	•	•	•	•	•	•	•	•	•	•
	F	•	•	•	•	•	1	1	2	3	4	4	4	3	2	1	1	•	•	•	•
	G	•	•	1	3	7	12	18	21	19	14	8	3	1	1	•	•	•	•	•	•
	H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	A	•	•	•	•	•	•	•	•	1	1	1	1	•	•	•	•	•	•	•	•
	B	•	•	•	•	•	•	1	2	5	8	11	11	13	11	8	4	2	1	•	•
	C	•	•	•	•	•	•	1	1	1	1	1	1	•	•	•	•	•	•	•	•
24	D	•	1	3	5	8	11	12	11	9	6	3	2	1	•	•	•	•	•	•	•
	E	•	•	•	1	2	2	3	3	2	1	1	•	•	•	•	•	•	•	•	•
	F	•	•	•	•	1	1	[illegible]	[illegible]	[illegible]	9	9	9	7	5	3	1	1	•	•	•
	G	•	•	2	5	12	23	34	40	[illegible]	26	15	6	2	1	•	•	•	•	•	•
	H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

* Centour Score < 0.5

† Groups are:

A. Radio Operators
B. Clerk-Typists
C. Control Tower Operators
D. Aircraft and Engine Mechanics
E. Radio Mechanics
F. Weather Observers
G. Airplane Sheet Metal Workers
H. Radar Mechanics

(Continued)

Table 15 (Cont.)

Discriminant Function 1(+20)

Discriminant Function 2(+20)	Group†	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	A	•	•	•	•	•	•	1	2	2	3	2	2	1	1	•	•	•	•	•	•
	B	•	•	•	•	•	1	2	5	9	16	23	27	26	22	14	8	4	2	1	•
	C	•	•	•	•	•	1	2	3	3	4	4	3	2	1	1	•	•	•	•	•
25	D	1	2	5	11	16	22	25	24	19	13	7	3	1	•	•	•	•	•	•	•
	E	•	•	1	2	4	6	7	7	5	4	2	1	•	•	•	•	•	•	•	•
	F	•	•	•	•	1	3	5	9	14	17	19	17	14	9	5	3	1	•	•	•
	G	•	1	3	[illegible]	19	37	55	64	57	42	24	10	4	1	•	•	•	•	•	•
	H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	A	•	•	•	•	1	1	3	4	6	7	7	6	3	2	1	•	•	•	•	•
	B	•	•	•	•	•	1	3	8	17	28	39	46	45	37	25	14	7	3	1	•
	C	•	•	•	•	1	2	4	7	9	10	10	8	5	3	2	1	•	•	•	•
26	D	2	4	9	18	28	39	46	42	34	23	13	6	2	1	•	•	•	•	•	•
	E	•	1	2	5	8	13	16	16	13	9	5	2	1	•	•	•	•	•	•	•
	F	•	•	•	1	2	5	10	17	24	31	32	30	24	16	9	5	2	1	•	•
	G	•	1	3	11	26	49	74	86	79	57	32	14	5	1	•	•	•	•	•	•
	H	•	•	1	1	1	•	•	1	1	1	•	•	•	•	•	•	•	•	•	•
	A	•	•	•	•	1	3	6	11	15	17	16	13	8	4	2	1	•	•	•	•
	B	•	•	•	•	1	2	5	14	25	43	59	69	67	54	36	20	9	4	1	•
	C	•	•	•	1	3	5	10	15	20	22	21	17	12	7	4	2	1	•	•	•
27	D	2	6	14	26	42	58	67	58	53	36	19	10	4	1	•	•	•	•	•	•
	E	•	1	3	8	16	24	31	32	26	18	10	4	2	•	•	•	•	•	•	•
	F	•	•	•	1	3	6	[illegible]	27	39	49	52	47	37	25	14	[illegible]	3	1	•	•
	G	•	1	4	12	29	56	84	99	91	66	37	17	5	•	•	•	•	•	•	•
	H	•	1	2	3	4	5	5	4	3	2	1	•	•	•	•	•	•	•	•	•
	A	•	•	•	1	3	6	13	22	30	35	33	26	16	9	4	1	•	•	•	•
	B	•	•	•	•	1	2	7	17	34	56	77	89	85	64	45	25	12	4	1	•
	C	•	•	1	2	5	10	19	29	38	42	40	32	22	13	7	3	1	•	•	•
28	D	3	7	17	33	54	75	87	85	70	49	28	14	6	2	1	•	•	•	•	•
	E	•	2	5	13	25	39	51	53	46	32	18	8	3	1	•	•	•	•	•	•
	F	•	•	1	2	5	12	23	38	55	68	72	65	50	33	19	9	4	1	•	•
	G	•	1	4	12	28	54	82	96	89	64	36	16	6	2	•	•	•	•	•	•
	H	1	2	4	7	10	14	15	13	10	6	3	1	•	•	•	•	•	•	•	•
	A	•	•	•	1	4	11	22	37	51	59	55	43	28	15	6	2	1	•	•	•
	B	•	•	•	•	1	3	8	20	38	63	87	99	94	75	49	27	13	5	2	•
	C	•	•	1	3	8	16	30	46	58	66	68	51	35	21	10	4	2	•	•	•
29	D	3	8	19	36	60	84	98	96	80	56	33	16	7	2	1	•	•	•	•	•
	E	•	2	7	17	33	54	72	77	67	48	28	13	5	2	•	•	•	•	•	•
	F	•	•	1	2	7	15	29	49	70	85	89	80	61	40	23	11	4	2	•	•
	G	•	1	3	10	23	45	67	79	73	[illegible]	[illegible]	14	5	1	•	•	•	•	•	•
	H	1	3	7	13	22	30	34	32	25	16	9	4	1	•	•	•	•	•	•	•
	A	•	•	1	2	6	15	31	52	72	83	78	61	39	21	9	3	1	•	•	•
	B	•	•	•	•	1	3	8	19	38	62	84	96	90	71	47	25	12	4	1	•
	C	•	•	1	4	10	22	40	61	79	88	84	67	46	27	14	6	2	1	•	•
30	D	3	7	18	34	57	80	95	94	79	56	33	16	7	2	1	•	•	•	•	•
	E	1	2	8	19	39	64	86	95	85	62	36	18	7	2	1	•	•	•	•	•
	F	•	•	1	3	8	17	33	55	79	95	99	88	66	43	24	11	5	2	•	•
	G	•	1	2	7	16	31	47	55	51	37	21	10	3	1	•	•	•	•	•	•
	H	1	4	9	20	34	50	60	60	49	34	19	9	3	1	•	•	•	•	•	•
	A	•	•	1	2	7	18	37	62	85	98	92	72	46	24	10	3	1	•	•	•
	B	•	•	•	•	1	3	7	17	32	52	71	80	75	58	38	21	9	4	1	•
	C	•	•	2	5	12	25	45	69	90	99	94	76	52	30	15	6	2	1	•	•
31	D	2	6	14	28	47	66	79	79	67	47	28	14	6	2	1	•	•	•	•	•
	E	1	2	7	18	38	64	89	100	91	68	41	20	8	3	1	•	•	•	•	•
	F	•	•	1	3	8	18	34	56	79	94	97	85	64	41	23	11	4	1	•	•
	G	•	•	1	4	9	18	28	35	31	22	13	6	2	1	•	•	•	•	•	•
	H	1	4	10	23	43	66	83	87	76	55	33	16	7	2	1	•	•	•	•	•
	A	•	•	1	2	7	18	36	61	85	97	95	71	46	24	11	4	1	•	•	•
	B	•	•	•	•	1	[illegible]	5	12	[illegible]	38	51	57	53	41	27	14	6	2	1	•
	C	•	•	2	5	11	24	43	66	86	95	89	72	49	29	14	6	2	1	•	•
32	D	1	4	10	19	33	47	56	57	48	35	19	11	5	2	1	•	•	•	•	•
	E	•	2	6	15	32	55	78	89	83	63	39	20	6	3	1	•	•	•	•	•
	F	•	•	1	3	7	16	31	50	70	33	85	74	55	35	19	9	4	1	•	•
	G	•	•	1	2	5	9	14	16	15	11	6	3	1	•	•	•	•	•	•	•
	H	1	3	9	21	41	67	90	100	92	70	45	23	10	4	1	•	•	•	•	•

(Continued)

Table 15 (Cont.)

Discriminant Function 1(+ 20)

Discriminant Function 2(+ 20)	Group†	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	A	•	•	1	2	6	15	30	51	71	81	76	59	38	20	9	3	1	•	•	•
	B	•	•	•	•	•	1	2	8	15	24	32	36	33	25	16	9	4	1	•	•
	C	•	•	1	4	9	20	35	53	69	76	[illegible]	57	39	23	11	5	2	1	•	•
33	D	1	2	6	12	20	28	35	35	30	22	13	7	3	1	•	•	•	•	•	•
	E	•	1	4	10	23	40	58	68	65	51	32	16	7	2	1	•	•	•	•	•
	F	•	•	1	2	6	13	25	40	55	65	66	57	42	26	14	7	3	1	•	•
	G	•	•	•	1	2	4	6	7	7	5	3	1	•	•	•	•	•	•	•	•
	H	1	2	6	15	31	53	76	89	87	70	47	26	12	5	1	•	•	•	•	•
	A	•	•	•	1	4	10	21	36	49	56	53	41	26	14	6	2	1	•	•	•
	B	•	•	•	•	•	1	2	4	8	13	18	19	18	13	9	5	2	1	•	•
	C	•	•	[illegible]	3	6	13	24	36	47	52	48	39	26	15	[illegible]	3	1	•	•	•
34	D	•	[illegible]	3	6	11	15	18	19	16	12	7	4	2	1	•	•	•	•	•	•
	E	•	1	2	6	14	25	37	48	44	35	22	12	5	2	•	•	•	•	•	•
	F	•	•	1	2	4	10	18	29	39	46	46	38	28	18	10	4	2	1	•	•
	G	•	•	•	•	1	1	2	3	2	2	1	•	•	•	•	•	•	•	•	•
	H	•	1	3	8	18	33	50	62	63	54	36	22	11	4	1	•	•	•	•	•
	A	•	•	•	1	2	6	12	21	29	33	31	24	15	8	4	1	•	•	•	•
	B	•	•	•	•	•	•	1	2	4	6	8	9	8	6	4	2	1	•	•	•
	C	•	•	1	2	4	8	14	21	27	30	28	22	15	9	4	2	1	•	•	•
35	D	•	•	1	2	5	7	8	9	8	6	3	2	1	•	•	•	•	•	•	•
	E	•	•	1	3	7	13	20	25	25	20	13	7	3	1	•	•	•	•	•	•
	F	•	•	•	1	3	6	11	18	24	28	28	24	17	11	6	3	1	•	•	•
	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	H	•	•	1	4	8	16	25	33	33	33	24	15	8	3	1	•	•	•	•	•
	A	•	•	•	•	1	3	6	10	14	16	15	12	7	4	2	•	•	•	•	•
	B	•	•	•	•	•	•	•	1	2	3	3	4	3	2	2	1	•	•	•	•
	C	•	•	•	1	2	4	7	10	13	14	13	11	7	4	2	1	•	•	•	•
36	D	•	•	•	1	2	3	3	3	3	2	1	1	•	•	•	•	•	•	•	•
	E	•	•	•	1	3	6	9	12	12	10	7	4	2	1	•	•	•	•	•	•
	F	•	•	•	1	2	4	6	10	14	16	15	13	9	6	3	1	1	•	•	•
	G	•	•	•	•	•	•	•	1	1	•	•	•	•	•	•	•	•	•	•	•
	H	•	•	•	1	3	6	10	14	16	15	12	8	6	2	1	•	•	•	•	•
	A	•	•	•	•	[illegible]	1	3	4	6	7	6	5	3	2	1	•	•	•	•	•
	B	•	•	•	•	•	•	•	•	•	•	1	1	1	1	1	•	•	•	•	•
	C	•	•	•	•	1	2	3	4	5	6	5	4	3	2	1	•	•	•	•	•
37	D	•	•	•	•	1	1	1	1	1	1	•	•	•	•	•	•	•	•	•	•
	E	•	•	•	•	•	2	4	4	5	4	3	2	1	•	•	•	•	•	•	•
	F	•	•	•	•	•	2	3	6	7	8	7	6	4	3	1	•	•	•	•	•
	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	H	•	•	•	•	1	2	3	5	6	6	5	3	2	1	•	•	•	•	•	•
	A	•	•	•	•	•	•	1	1	2	2	2	2	1	1	•	•	•	•	•	•
	B	•	•	•	•	•	•	•	•	1	1	•	•	•	•	•	•	•	•	•	•
	C	•	•	•	•	•	1	1	1	2	2	2	1	1	1	•	•	•	•	•	•
38	D	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	E	•	•	•	•	1	1	1	2	2	2	1	1	•	•	•	•	•	•	•	•
	F	•	•	•	•	•	1	1	2	3	3	3	2	2	1	1	•	•	•	•	•
	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	H	•	•	•	•	•	•	1	1	1	2	1	1	1	•	•	•	•	•	•	•
	A	•	•	•	•	•	•	•	•	1	1	•	•	•	•	•	•	•	•	•	•
	B	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	C	•	•	•	•	•	•	•	•	1	1	•	•	•	•	•	•	•	•	•	•
39	D	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	E	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	F	•	•	•	•	•	1	1	1	1	1	•	•	•	•	•	•	•	•	•	•
	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

TABLE 16

REGIONS IN WHICH CENTOUR SCORES OF EACH GROUP ARE 10 OR MORE

Discriminant Function 1

Discriminant Function 2

	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
36							- - - - - - -h	a- c- ef -h	a- c- ef -h	a- c- ef -h	a- c- -f -h	a- c- -f - -						
35						- - - - e- -h	a- c- ef -h	a- c- ef -h	a- c- ef -h	a- c- ef -h	a- c- ef -h	a- c- -f -h	a- c- -f - -	- - - - -f - -				
34					- - -d e- -h	a- cd ef -h	a- cd ef -h	a- cd ef -h	a- cd ef -h	ab cd ef -h	ab c- ef -h	ab c- ef -h	ab c- -f -h	ab c- -f - -	- - - - -f - -			
33				- - -d e- -h	- - -d e- -h	a- cd ef -h	a- cd ef -h	a- cd ef -h	ab cd ef -h	ab cd ef -h	ab cd ef -h	ab c- ef -h	ab c- -f -h	ab c- -f - -	-b c- -f - -			
32			- - -d - - - -	- - -d e- -h	- - cd e- -h	a- cd ef -h	a- -d ef gh	ab cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef -h	ab cd ef -h	ab c- -f -h	ab c- -f - -	ab c- -f - -	-b - - - - - -		
31			- - -d - - -h	- - -d e- -h	- - cd e- -h	a- cd ef gh	a- cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef -h	ab c- -f - -	ab c- -f - -	ab c- -f - -	-b - - -f - -		
30			- - -d - - - -	- - -d e- -h	- - cd e- gh	a- cd ef gh	a- cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef g-	ab c- -f - -	ab c- -f - -	-b c- -f - -	-b - - -f - -	-b - - - - - -	
29			- - -d - - - -	- - -d e- gh	- - -d e- gh	a- cd ef gh	a- cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef g-	ab cd ef g-	ab c- -f - -	-b c- -f - -	-b c- -f - -	-b - - -f - -	-b - - - - - -	
28			- - -d - - - -	- - -d e- g-	- - -d e- gh	- - cd ef gh	a- cd ef gh	ab cd ef gh	ab cd ef gh	ab cd ef g-	ab cd ef g-	ab cd -f g-	ab c- -f - -	-b c- -f - -	-b - - -f - -	-b - - - - - -	-b - - - - - -	
27			- - -d - - - -	- - -d - - g-	- - -d e- g-	- - -d e- g-	- - cd ef g-	ab cd ef g-	ab cd ef g-	ab cd ef g-	ab cd ef g-	ab cd -f g-	-b c- -f - -	-b - - -f - -	-b - - -f - -	-b - - - - - -		
26				- - -d - - g-	- - -d - g-	- - -d e- g-	- - -d ef g-	- - -d ef g-	-b -d ef g-	-b cd -f g-	-b cd -f g-	-b - - -f g-	-b - - -f - -	-b - - -f - -	-b - - - - - -	-b - - - - - -		
25				- - -d - - - -	- - -d - - g-	- - -d - - g-	- - -d - - g-	- - -d - - g-	- - -d -f g-	-b -d -f g-	-b - - -f g-	-b - - -f g-	-b - - -f - -	-b - - - - - -	-b - - - - - -			
24					- - - - - - g-	- - -d - - g-	- - -d - - g-	- - -d - - g-	- - - - - - g-	- - - - - - g-	-b - - - - g-	-b - - - - - -	-b - - - - - -	-b - - - - - -				
23						- - - - - - g-	- - - - - - g-	- - - - - - g-	- - - - - - g-	- - - - - - g-								
22								- - - - - - g-										

a = Radio Operators
b = Clerk-Typists
c = Control Tower Operators
d = Aircraft and Engine Mechanics
e = Radio Mechanics
f = Weather Observers
g = Airplane Sheet Metal Workers
h = Radar Mechanics

maximum of an airman's eight centour scores, relatively few airmen will be assigned to the control tower operator, radio operator, and radio mechanic specialties. It would seem that the better procedure would be to rule out certain specialties associated with a particular pair of discriminant scores and then to assign to remaining specialties in terms of both the airman's choices and quotas set by the requirements of the service.

The interpretation of the discriminants may be facilitated by working within a fixed error allowance. If this error is chosen at 10 per cent for every group, Table 16 provides the discriminant information for guidance couseling of airmen concerning the eight Air Force specialties studied. In this table the groups where the centour scores exceed 9 are identified for each pair of discriminant scores. With this table it is possible to determine immediately all the groups in which an airman may belong,[12] once his two discriminant scores are known. If a dash (—) appears in place of the letter for any group, it indicates that less than 10 per cent of the airmen in this group made discriminant scores outside the boundary set by the particular pair of discriminant scores.

We considered earlier the example of an airman whose discriminant scores were 19 and 24. If we look up the "19" column of Table 16 in "24" row, we find the letters b and g, indicating that, as we saw before, he has centour scores of 10 or more in the Clerk-Typist group and in the Sheet Metal group.

A table similar to 16 could be made up for a fixed error allowance of 15 per cent (instead of 10 per cent) and it would show only a "g" in the 19,24 cell.

If the 17 tests were discriminating perfectly among the eight groups, we would find only a single group-letter in each cell of the table. It is clear from Table 16 that this perfection is not achieved in the data studied. Some indication of how far the Airman Classification Battery falls short of discriminating perfectly among the eight groups is found in Table 16 by reference to the central region enclosed in solid lines. In this region of 16 pairs of discriminant scores, the centour scores for all eight groups exceed 9. A large fraction of airmen obtain one or another of these 16 pairs of discriminant scores.

[12] Subject to the fixed error percentage, of course.

C. USE OF IBM MACHINES FOR REPORTING CENTOUR SCORES

A master deck of centour scores associated with each pair of discriminant scores could be prepared. After discriminant scores are computed for each airman by the method described in Appendix D of the complete report (10), the two decks of cards could be sorted together and centour scores reproduced into the detail cards for each airman. A tabulated roster of centour scores would provide a guidance counselor with almost all the information that the Airman Classification Battery contains concerning prediction of group identity.

EFFICIENCY OF THE AIRMAN CLASSIFICATION BATTERY FOR GUIDANCE CONCERNING THESE SPECIALTIES

A. THE PROBLEM OF GROUP MEMBERSHIP

Questioners of discriminant analysis have doubted that the technique is appropriate for the problems of personnel classification and guidance. They propose that the technique is appropriate for a field like taxonomy, where group identity is ascertained with considerable certainty and where an object or individual can belong to only one group. The questioners doubt that the establishment of original group identification in personnel selection and guidance meets these conditions.

As an instance, suppose one had a set of skulls from some tribe where the sex of the former owners of the skulls was known in each case. Now suppose a new skull were found and the sex of the former owner were not known. Can the sex of the former owner of this skull be determined by measurements of the skull? Discriminant analysis is highly appropriate for this problem, it is agreed. But the utility of discriminant analysis for personnel classification is still questioned by some.

In an organization like the Air Force and in a society like that of the United States personnel classification does and must take place. The Air Force and non-military American societies have progressed to the point where, for all practical purposes, each working member of the society performs only one job. The worker or airman gets classified as a radio mechanic, clerk-typist, weather observer, or some such thing. If either of these men is asked what he

does his response is likely to be in terms of a single occupation.

Thus, the facts are that occupations do exist, that rules exist for classifying people into occupations, that most individuals may be classified into only one occupation, and that vocational counselors believe they know the abilities required by each occupation. In this situation two questions may be examined. Either one may believe that one is studying all the variates required for differentiation of occupations and ask if there is any real basis for thinking that certain occupations require different abilities, or one may believe that occupations are different and ask if a certain set of variates are the ones in terms of which the occupations are different. Discriminant analysis is appropriate to the investigation of either of these questions.

Actually, this study met neither of these conditions exactly. However, it probably resembles the first more than the second. The Airman Classification Battery represents an extensive, but not exhaustive, list of variates. The problem involves study of how extensively these variates figured in the definition of the specialties.

As of 1 July 1950, there were some 6105 airmen who had taken all the tests of Airman Classification Battery AC-1 after 15 November 1948 and who were then known as specialists in one of eight Air Force specialties because they had satisfactorily completed training in one of them. These airmen got to be the particular kinds of specialist they were because of ways in which they, Indoctrination Wing counselors, and assignment officials combined information concerning themselves and the Air Force. Discriminant analysis provides a means of studying how the Airman Classification Battery part of this information was combined. No case is made for the point that these combinations and the sectioning of the space defined by the combinations represent inalienable definitions of the aptitudes required for each of these eight specialties, nor is any case made for the point that an airman is unable to perform at more than one specialty. However, the points are made with considerable emphasis that the discriminant functions of this study represent the way in which these 6105 airmen and their Indoctrination Wing Counselors sectioned the 17-dimensional Airman-Classification-Battery-space when faced with the problem of choosing only one of eight specialties, and that this knowledge, if applied to new airmen, would provide a means of maintaining this orientation.

B. ORIENTATION OF THE DISCRIMINANT FUNCTIONS

The most striking finding of this study has been that all the essential information concerning separation of 17-dimensional centroids of eight groups as diverse as clerk-typists, sheet metal workers, aircraft and engine mechanics, radar mechanics, and weather observers was condensed to a two-dimensional space by the airmen and Indoctrination Wing counselors. Kelley (4, pp. 33–34) as early as 1935 suggested that this might be the general order of magnitude of understanding of similar situations.

Some idea of the amount of attention given to each of the 17 variates is obtained from study of Table 2. The largest positive coefficients of Discriminant Function 1 are for clerical background, word knowledge, and numerical operations, while the largest negative coefficients are for mechanical background, aviation information, and electrical information. Large positive coefficients of Discriminant Function 2 are for radio operator background, electrical information, memory for landmarks, dial and table reading, and numerical operations, while large negative coefficients are associated with tool functions, and craftsman, equipment operator, and radio operator backgrounds.

The manner in which scores of individual airmen are distributed in this two-dimensional discriminant space has been shown in Figure 5.9 (Appendix C). However, to simplify matters for this discussion, only group centroids are shown in Figure 7.1 (Appendix C). This chart and the discriminant function coefficients suggest that Discriminant Function 1 reflects mechanical ability. If this scale is reversed, a procedure which has no effect on separation of the group centroids, we see that the specialties are arranged in the following order from high to low:

Aircraft and engine mechanic
Radio mechanic, and radar mechanic
Sheet metal worker
Control tower operator, and radio operator
Weather observer
Clerk-typist.

These specialties are quite clearly arranged on a mechanical vs. non-mechanical continuum.

Those specialties near the top require more mechanical ability than those near the bottom. Clerk-typists require little mechanical ability, if any.

The specialties are arranged from high to low on Discriminant Function 2 as follows:

Radar mechanic
Radio operator, and control tower operator
Radio mechanic
Weather observer
Aircraft and engine mechanic, and clerk-typist
Sheet metal worker.

This order suggests an intellectual dimension. The radar mechanic, radio operator, and control tower operator specialties require more intellect than the aircraft and engine mechanic, clerk-typist, and sheet metal worker specialties.

C. EFFICIENCY OF THE DISCRIMINANT FUNCTIONS

The size of the latent roots reported on page 5 and the overlap of group distributions demonstrated in Figure 5.9 (Appendix C) and Tables 15 and 16 provide evidence of the rather small amount of group separation achieved by use of the Airman Classification Battery for assignment of airmen to the eight specialties of this study. A fairly considerable scatter of many abilities is represented by airmen who have satisfactorily completed training in each of the eight specialties.

D. RESEARCH IMPLICATIONS

Air Force and civilian psychologists have spent considerable energy in the development of tests which are related to degrees of success in various jobs. The pattern for this type of research was established as early as World War I. Of course, the rather extensive utilization of the factor analysis concept after 1930 has altered the pattern in the sense of suggesting different tests that could be used in investigating relationships between degrees of competence within a job, but it has not altered the goal of eliminating tests that are not related to the within-job competence measures.

Preoccupation with this purpose and the lack of a proper method of analysis have caused psychologists to ignore the situation with which a vocational counselor must deal. The grossest and most important responsibility of the vocational counselor is not that of judging success on a given job but that of judging for which of a host of jobs the client has the requisite aptitudes. Discharge of this responsibility involves the assumptions that there are a number of abilities of humans and that different degrees and combinations of them are required for different jobs. Study of degrees of success within each type of job furnishes no information about comparison of jobs in terms of abilities required. Research which is completely oriented in this first manner, as it almost always has been, may have resulted in a selection of variates entirely different from the ones appropriate for differentiating the ability requirements of various jobs. Therefore, a thorough examination of the psychometric literature needs to be made in order to introduce for reconsideration those variates which have been abandoned because they have little relationship to degree of success within a job, but which may provide important information for comparison of different jobs.

When these variates have been recaptured, and when it has been established that people in different jobs perform differently on variates used frequently at present, discriminant analyses need to be made with more variates and more jobs than those of this study in order to expand understanding of how these variates may be combined for maximum job distinction with a minimum of variates.

Several kinds of studies could be made to determine the reality of group identity. One type of study would be to afford tryout experiences in each specialty for each airman in order to find the specialty which is considered most appropriate. Research of this nature would be expensive, of course, but it would probably be the most valid. A second type of study would be to assign airmen to specialties in a manner such that group homogeneity of discriminant scores were increased and to study changes in the satisfaction both of the schools with the various groups and of the airmen with their assignments.

Definition of job families by means of discriminant analysis promises to be a much sounder procedure than definition in terms of factor structure. Finding the members of one specialty group close by the members of another in the discriminant space suggests that so far as the tests can tell these two specialty groups are indistinguishable from each other. A factor analysis of each

group might yield the same factor loadings in one as in the other, even though the tests could readily distinguish the two groups. If the tests can distinguish the two groups, they will not appear indistinguishable in the discriminant space. The present study suggests strongly that radio operator, control tower operator, and weather observer specialties are jobs belonging to the same job family, as judged by the 17 tests and in the setting in which the tests were used. Expansion of discriminant analysis research will provide a much more realistic foundation for development of job families than exists at present.

In this connection it might be noted that multiple discriminant analysis and multiple regression analysis represent, in certain respects, complementary techniques.

SUMMARY

In order to provide valid information for the guidance of individuals who have not yet selected an occupation it is necessary that three questions be investigated. These are: (a) Do the distributions of the several occupations occupy different regions of the multivariate space? (b) If so, may the distinctions among occupations be described by a space of fewer dimensions than the original one? and (c) If so, can probability statements be derived which will indicate the probabilities with which a given point in the reduced space could have been drawn from the distribution of each occupation? The multiple discriminant functions obtained from Joseph G. Bryan's generalization (1) of R. A. Fisher's two-group discriminant function provide a means of answering all three of these questions. Bryan's method defines not more than G-1 discriminant functions when the number of groups, G, is less than the number of variates.

A study was made of 6105 airmen who, prior to 1 July 1950, had satisfactorily completed training in one of eight Air Force specialties. These airmen were classified into such diverse specialties as airplane sheet metal workers, clerk-typists, aircraft and engine mechanics, and radar mechanics. (A complete list of specialties is given on page 4).

Each airman completed all tests of Airman Classification Battery AC-1. The airman chose his specialty in consultation with an Indoctrination Wing counselor who had access to the aptitude indexes and test scores of the Airman Classification Battery.

Multiple discriminant analysis of these data revealed that essentially all the information concerning separation of the eight-specialty centroids in the 17-dimensional-Airman-Classification-Battery space is described by two linear combinations of the 17 variates. The remaining five among-means-of-groups-to-within-groups ratios were negligible in comparison to the first two. The two functions seem to represent mechanical ability and intellectual ability.

Two discriminant scores were computed for each airman. Study of distributions of the two scores for each specialty revealed that an assumption of bivariate normality for the discriminant score distributions was tenable for all but the aircraft and engine mechanic specialty. Despite this one exception, centour scores were computed for each group upon the assumption that the bivariate discriminant score distributions of each group are normal. A centour score is a function of the probability that a given pair of discriminant scores comes from a particular specialty. These scores would be exceptionally helpful for career guidance of new airmen except for the large overlap of the eight groups in the Airman-Classification-Battery space.

Manuscript received 10 June 1951

BIBLIOGRAPHY

1. BRYAN, J.G. A method for the exact determination of the characteristic equation and latent vectors of a matrix with applications to the discriminant function for more than two groups. Unpublished Doctoral dissertation, Harvard Univ., 1950.

2. FISHER, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, VII, 1936, 179–188.

3. FISHER, R.A. The statistical utilization of multiple measurements, *Ann. Eugen.*, VIII, 1938, 376–386.

4. FISHER, R.A. *Statistical methods for research workers.* (10th Ed.) London: Oliver and Boyd, 1946.

5. GRAGG, D.B., and GORDON, MARY AGNES. *Validity of the Airman Classification Battery AC-1.* San Antonio, Tex.: Human Resources Research Center, Lackland Air Force Base, December 1950. (*Research Bulletin* 50–3.)

6. KELLEY, T.L. *Essential traits of mental life.* Cambridge, Mass.: Harvard University Press, 1935.

7. PEARSON, K. *Tables for statisticians and biometricians, Part I.* (2nd Ed.) University College, London: The Biometric Laboratory, 1924.

8. Personnel Research Laboratory. *Research Planning Conference on objective measurement of motivation and temperament.* San Antonio, Tex.: Human Resources Research Center, Lackland Air Force Base, June 1951. (*Conference Report* 51–?.)

9. RIETZ, H.L. *Mathematical statistics.* (4th Print.) John S. Swift Co., Inc., 1943.

10. TIEDEMAN, D.V., BRYAN, J.G., and RULON, P.J. *The utility of the Airman Classification Battery for Assignment of airmen to eight Air Force specialties.* Cambridge, Mass.: Educational Research Corporation, June 1951.

APPENDIXES

APPENDIX A

MULTIPLE DISCRIMINANT ANALYSIS*

1. GENERAL DESCRIPTION

Let X_{pgj} denote the value of the j^{th} variate ($j = 1, 2, \ldots, n$) of the p^{th} individual ($p = 1, 2, \ldots, N_g$) in the g^{th} group ($g = 1, 2, \ldots, G$) where $G < n$. Without loss of generality we may define the X_{pgj} so that

$$\sum_g \sum_p X_{pgj} = 0 \qquad [2.1]$$

for all values of j. In general, separate group means such as

$$\overline{X}_{gj} = \frac{1}{N_g} \sum_p X_{pgj} \qquad [2.2]$$

do not vanish.

A linear function of the X_j such as

$$y = v_1X_1 + v_2X_2 + \ldots + v_nX_n \qquad [2.3]$$

would have group means of

$$\bar{y}_g = \frac{1}{N_g} \sum_p y_{pg} \qquad [2.4]$$

where y_{pg} is the value of the linear function for the p^{th} individual in the g^{th} group. Because of the restrictions defined by equation [2.1], the among means of groups sum of squares of y is

$$\sum_g N_g \bar{y}_g^{\,2}. \qquad [2.5]$$

The within groups sum of squares is

$$\sum_g \sum_p (y_{pg} - \bar{y}_g)^2. \qquad [2.6]$$

The coefficients, $v_1, v_2, \ldots, v_n$ of y are defined by Fisher's (2, pp. 179–188) criterion of discriminant analysis, i.e. by maximization of the ratio λ where

$$\lambda = \frac{\sum_g N_g \bar{y}_g^{\,2}}{\sum_g \sum_p (y_{pg} - \bar{y}_g)^2} \qquad [2.7]$$

The function λ has several extrema each of which is indicative of a distinct dimension of the subspace defined by the group means. All discriminant functions are obtained from the same initial ratio λ (1, pp. 132–138).

Maximization of λ is accomplished by a combination of two effects, increase in the among means of groups sum of squares and decrease in the within groups sum of squares. Therefore, the linear functions defined by the discriminant analysis criterion provide as much separation of group centroids as is possible within the restriction of having the groups as homogeneous as possible. The utility for group classification of functions with this property is obvious.

Manipulation of the quadratic forms of the variates required for discriminant analysis is facilitated by use of matrix algebra. Therefore let us define the symmetrical matrices

$$A = \|a_{ij}\|, \; a_{ij} = \sum_g N_g \overline{X}_{gi} \overline{X}_{gj} = a_{ji}, \; (i,j = 1,2,\ldots,n); \qquad [2.8]$$

$$W = \|w_{ij}\|, \; w_{ij} = \sum_g \sum_p (X_{pgi} - \overline{X}_{gi})(X_{pgj} - \overline{X}_{gj}) = w_{ji}, \; (i,j = 1,2,\ldots, n); \qquad [2.9]$$

and the column vector

$$v = [v_j], \; (j = 1,2, \ldots, n). \qquad [2.10]$$

In this notation equations [2.5], [2.6], and [2.7] may be written as

$$\sum_g N_g \bar{y}_g^{\,2} = v'Av, \qquad [2.5]$$

*This appendix is a reproduction of Chapter 2 of the complete report (10). It is labelled "Appendix A" in this condensation.

$$\sum_g \sum_p (y_{pg} - \bar{y}_g)^2 = v'Wv, \qquad [2.6]$$

and

$$\lambda = \frac{v'Av}{v'Wv} \qquad [2.7]$$

It has been shown (1, pp. 131–132) that the matrix equation defining the vector v is

$$(v'Wv)\,Av - (v'Av)\,Wv = 0.$$

This may be rewritten as

$$(A - \lambda W)\,v = 0.$$

Defining

$$R = W^{-1}A,$$

this becomes

$$(R - \lambda I)\,v = 0 \qquad [2.11]$$

where I is the unit matrix.

The coefficients of the discriminant functions are determined by the latent vectors of R, and the corresponding latent roots of R equal the respective ratios of among-groups to within-groups sums of squares. By considering the rank of the matrix A, it is a simple matter to show that the number of solutions of [2.11] such that $\lambda \neq 0$ is at most equal to the smaller of the two integers G−1, and n. Consequently, letting r stand for the smaller number, the total discriminative power of the variates is exhausted by r linear functions defined in the manner stated. Among these, all functions corresponding to distinct values of λ are uncorrelated as they stand. Repeated roots other than zero are possible but unlikely to occur. If one or more multiple roots should occur, however, the vectors corresponding to any one of them are already uncorrelated with the vectors corresponding to all different roots and can be chosen in such a way as to be uncorrelated among themselves. The numerical values of these functions are independent of the origin of coordinates, the units of measurement, and, in fact, independent of any non-singular linear transformation of the variates (1, pp. 138–139).

II. GENERALIZED COMPUTATIONAL PROCEDURE

Bryan (1, pp. 139–147) has demonstrated that the latent roots and vectors of the matrix R may be computed in the following manner:

1. Compute elements of matrices A and W from the original data.
2. Partition A into four parts:

$$A = \begin{Vmatrix} A_{11} & A_{12} \\ & \\ A_{12}' & A_{22} \end{Vmatrix}$$

where A_{11} is r×r, A_{12} is r×(n−r), and A_{22} is (n−r)×(n−r). The rank r of A cannot exceed the smaller of the two integers G−1 and n. Now if n is substantially greater than G−1 (as was the case here), considerable computational saving can be obtained by condensing R into a matrix of lower order having r rows and columns. For simplicity it has been assumed that r = G−1, but this is not crucial, for if r<G−1 (mathematically possible but statistically improbable if n≥G−1) the working directions still apply except that r must be interpreted as the true rank and A_{11} must consist of r linearly independent rows and columns. Bryan has shown that the latent roots of the condensed matrix are the same as the non-vanishing roots of R, and the corresponding latent vectors are simply related to those of R (1, pp. 142–147).

3. Compute the matrix $B = A_{11}^{-1} A_{12}$ by the modified Crout method described in Appendix C of the complete report (10). Set up the initial matrix M thus.

$$M = \begin{Vmatrix} A_{11} & A_{12} \\ & \\ I_r & 0 \end{Vmatrix}$$

When the auxiliary matrix is computed according to the modified rules, B appears as the lower right hand panel (set off by dotted lines). The matrix B is used to condense R. This is possible because the fact that the rank r of A is less than n implies linear dependence such that

$$A_{12} = A_{11}B, \quad A_{22} = B'A_{11}B$$

4. Write four transformation matrices as follows:

$$J = \left\| \begin{matrix} I_r & 0_{12} \\ -B' & I_{n-r} \end{matrix} \right\|$$

$$J' = \left\| \begin{matrix} I_r & -B \\ 0'_{12} & I_{n-r} \end{matrix} \right\|$$

$$H = \left\| \begin{matrix} I_r & 0_{12} \\ B' & I_{n-r} \end{matrix} \right\|$$

$$H' = \left\| \begin{matrix} I_r & B \\ 0'_{12} & I_{n-r} \end{matrix} \right\|$$

$$H^* = \left\| \begin{matrix} I_r \\ B' \end{matrix} \right\|$$

5. Derive an intermediate matrix $P^* = H' W^{-1} H^*$. P^* emerges as the lower right hand panel of the matrix when the modified Crout auxiliary method is applied to the matrix

$$M = \left\| \begin{matrix} W & H^* \\ H' & 0 \end{matrix} \right\|$$

6. Compute $Q^* = P^* A_{11}$ and partition Q^* into the parts

$$\left\| \begin{matrix} Q_{11} \\ Q_{21} \end{matrix} \right\|$$

7. Compute the characteristic equation and latent vectors of Q_{11} as follows:

Starting with a two-element row vector f_1, the elements of which are always known in advance, one generates a matrix-vector sequence B_2, f_2, B_3, f_3, ..., B_n, f_n by matrix multiplication and addition. The elements of the final vector f_n are the coefficients of the characteristic equation, while matrix B_n immediately preceding it furnishes the latent vectors. We shall define the initial vector f_1 presently; for the moment let us see how one member of the sequence is evolved from its predecessors. Throughout this subsection we shall observe the convention that r, s, t, denote consecutive integers, that is, $r = s - 1$, $t = s + 1$.

Let A_s denote the submatrix comprising the first s rows and columns of the nxn matrix $A = \|a_{ij}\|$. Partition A_s into two panels: one denoted by α_s and including its first r rows, the other denoted by a_s and consisting of its last row.
Symbolically,

$$A_s = \|a_{ij}\| \qquad i,j = 1,2,\ldots,s$$

or, in partitioned form,

$$A_s = \left\| \begin{matrix} \alpha_s \\ a_s \end{matrix} \right\| , \quad a_s = (a_{sj})$$

Now by the definition we choose to adopt, the row vectors f_k always have k+1 elements, so that f_r has s elements f_{rj}, $j = 1,2,\ldots,s$.

We now state the rule by which B_s and f_s ($s \geq 2$) are found. The last *row* of B_s is simply f_r: the first *column* of B_s is filled in by erecting a vertical array of r zero elements above the first element of f_r. Let ${}^s b_j$ denote the j th column of B_s. The bottom element of each column, being an element of f_r, is already given. We need therefore determine only the vertical array of r elements above f_r. This vertical array is found by the recurrence relation

First r elements of ${}^s b_j = \alpha_s \, {}^s b_{j-1}$, $j = 2,3,\ldots,s$
First r elements of ${}^s b_1 = 0$

Introducing the sx1 null vector ${}^s b_0 = 0$ for all s, we arrive at f_s by the rule

$$f_{sj} = a_s {}^s b_{j-1} - f_{rj} \qquad j = 1,2,\ldots,s$$

$$f_{st} = a_s {}^s b_s$$

It remains to state the definition of f_1. This is

$$f_1 = (-1 \quad a_{11})$$

The definition f_a is precise as stated. For those who prefer completely symbolic definitions, the rule for B_a is

$$^a b_j = \begin{bmatrix} a_a\,{}^a b_{j-1} \\ \\ \\ f_{rj} \end{bmatrix} \qquad j = 1, 2, \ldots, n$$

where $^a b_0 = 0$ (1, pp. 80–81).

8. Obtain the latent roots of the characteristic equation by the Birge-Vieta process.

9. Let the latent roots of Q_{11} be $\lambda_j,\ j = 1, 2, \ldots, r$ and define the $r \times r$ matrix

$$L = \left\| \lambda_j^{\,i-1} \right\| \qquad i, j = 1, 2, \ldots, r$$

where i stands for the row index.

10. Premultiply L by B_r and call this result S^*.

11. Compute $J'Q^*S^*$ and conventionalize the result. The columns of this matrix give the coefficients of the r discriminant functions.

This routine is illustrated in Sections C and E in this appendix and by the major example on pages 10–17. Detailed computing instructions and work sheets are given in Appendix C of the complete report (10).

III. ILLUSTRATION OF DISCRIMINANT ANALYSIS—TWO GROUPS AND THREE VARIATES

The scores of clerk-typists (course number 40501) and radar mechanics (course number 77501)[1] on the mechanical key of the Biographical Inventory (X_2), the Electrical Information test (X_{12}), and the Tool Function test (X_{15}) were selected to illustrate discriminant function computations in this simplest case, two groups and three variates. These groups and variates were chosen because the data showed considerable promise of providing discrimination.

The case of two groups is most easily handled by Fisher's multiple regression method, but as an introductory example of the general technique it seemed desirable to keep the numerical work at a minimum. It happens that some of the computational steps unavoidably appear roundabout or unnecessary. This is due solely to the trivial nature of the two-group case.

The discriminant analysis routine is as follows:

1. *Compute Sums of Squares and Cross Products* $(\sum_p X_{pgi}X_{pgj})$ *for Each Group*

By progressive digiting of the IBM cards compute the sums of squares and cross products of the variates $(\sum_p X_{pgi}X_{pgj})$. The values are as follows:

		Variate		
	Group	2	12	15
2	B	41327	47813	41359
	H	4833	5481	4842
	Sum	46160	53294	46201
12	B	47813	[illegible]	54435
	H	5481	6529	5736
	Sum	53294	71509	60171
15	B	41359	54435	50984
	H	4842	5736	5245
	Sum	46201	60171	56229

2. *Determine the Sum of Scores* $(\sum_p X_{pgj})$ *for Each Group*

In progressive digiting, the final sum of each run of the IBM cards gives the sum of scores $(\sum_p X_{pgj})$. The values are:

	Variate			
	2	12	15	N_g
B	8337	10806	9286	1966
H	677	797	701	99
Sum	9014	11603	9987	2065

3. *Compute Matrix A*

For each group, compute $\dfrac{(\sum_p X_{pgi})(\sum_p X_{pgj})}{N_g}$

for all combinations of i and j.

Sum these products for all groups getting

$$r_{ij} = \sum_g \left[\frac{(\sum_p X_{pgi})(\sum_p X_{pgj})}{N_g} \right]$$

Compute

$$s_{ij} = \frac{(\sum_g \sum_p X_{pgi})(\sum_g \sum_p X_{pgj})}{\sum_g N_g}$$

[1] Clerk-Typists were assigned EMC Code B and radar mechanics EMC Code H.

The elements a_{ij} of the A matrix are then defined by

$$a_{ij} = r_{ij} - s_{ij}$$

The results of these computations are as follows:

		Variate		
		2	12	15
	B	35253.7991	45823.8159	39378.1190
	H	4629.5859	5450.1919	4793.7071
2	r_{1j}	39983.3850	51274.0078	44171.8261
	s_{1j}	39347.3104	50648.6402	43594.5850
	$r_{1j}-s_{1j}$	636.0746	625.3676	577.2411
	B	45823.8159	59394.5249	51039.9369
	H	5450.1919	6416.2525	5643.4040
12	r_{2j}	51274.0078	65810.7774	56683.3409
	s_{2j}	50648.6402	65195.9366	56115.8165
	$r_{2j}-s_{2j}$	625.3676	614.8408	567.5244
	B	39378.1190	51039.9369	43860.5270
	H	4793.7071	5643.4040	4963.6465
15	r_{3j}	44171.8261	56683.3409	48824.1735
	s_{3j}	43594.5850	56115.8165	48300.3240
	$r_{3j}-s_{3j}$	577.2411	567.5244	523.8495

For convenience matrix A is recopied here:

	Variate			
	2	12	15	Sum
2	636.0746	625.3676	577.2411	1838.6833
12	625.3676	614.8408	567.5244	1807.7328
15	577.2411	567.5244	523.8495	1668.6150
Sum	1838.6833	1807.7328	1668.6150	5315.0311

4. *Compute Matrix W*

For each group compute and record

$$\frac{N_g \sum_p X_{pgi} X_{pgj} - (\sum_p X_{pgi}) (\sum_p X_{pgj})}{N_g}$$

for all pairs of i and j. In each element sum over the groups. This element is w_{ij} of the W matrix. Results are as follows:

		Variate		
		2	12	15
	B	5973.2009	1989.1841	1980.8810
2	H	203.4141	30.8081	48.2929
	Sum	6176.6150	2019.9922	2029.1739
	B	1989.1841	5585.4751	3395.0631
12	H	30.8081	112.7475	92.5960
	Sum	2019.9922	5698.2226	3487.6591
	B	1980.8810	3395.0631	7123.4730
15	H	48.2929	92.5960	281.3535
	Sum	2029.1739	3487.6591	7404.8265

For convenience matrix W is recopied here:

	Variate			
	2	12	15	Sum
2	6176.6150	2019.9922	2029.1739	10225.7811
12	2019.9922	5698.2226	3487.6591	11205.8739
15	2029.1739	3487.6591	7404.8265	12921.6595
Sum	10225.7811	11205.8739	12921.6595	34353.3145

5. *Compute Matrix T as a Check*

Compute and record

$$t_{ij} = \frac{(\sum_g N_g) (\sum_g \sum_p X_{pgi} X_{pgj}) - (\sum_g \sum_p X_{pgi}) (\sum_g \sum_p X_{pgj})}{\sum_g N_g}$$

for all combinations of i and j. Add and record a_{ij} and w_{ij} from the A and W matrices. As a check $t_{ij} = a_{ij} + w_{ij}$. Results are as follows:

		Variate		
		2	12	15
2	t_{ij}	6812.6896	2645.3598	2606.4150
	$a_{ij}+w_{ij}$	6812.6896	2645.3598	2606.4150
12	t_{ij}	2645.3598	6313.0634	4055.1835
	$a_{ij}+w_{ij}$	2645.3598	6313.0634	4055.1835
15	t_{ij}	2606.4150	4055.1835	7928.6760
	$a_{ij}+w_{ij}$	2606.4150	4055.1835	7928.6760

6. *Compute Matrix B*

Since $r = 1$, the submatrices A_{11} and A_{12} are

$$A_{11} = \begin{Vmatrix} 636.0746 \end{Vmatrix}$$

and

$$A_{12} = \begin{Vmatrix} 625.3676 & 577.2411 \end{Vmatrix}$$

The initial matrix for the Crout computational method of determining $B = A_{11}^{-1} A_{12}$ is:

$$\begin{Vmatrix} 636.0746 & 625.3676 & 577.2411 \\ 1 & 0 & 0 \end{Vmatrix}$$

The Crout computational method[2] yields the following modified auxiliary matrix:

			Sum and Check
636.0746	0.983167069	0.907505346	2.890672416 1.890672415
1	0.983167069	0.907505346	1.890672416 1.890672415

[2] See Appendix C of complete report (10).

of which the lower right hand corner is B or

$$B = (0.983167069 \quad 0.907505346)$$

Therefore

$$B' = \begin{bmatrix} 0.983167069 \\ 0.907505346 \end{bmatrix}$$

7. *Write the Transformation Matrices J, J', H, H', and H**

In accordance with the definitions of Section B,

$$J = \begin{Vmatrix} 1 & 0 & 0 \\ -0.983167069 & 1 & 0 \\ -0.907505346 & 0 & 1 \end{Vmatrix}$$

$$J' = \begin{Vmatrix} 1 & -0.983167069 & -0.907505346 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

$$H = \begin{Vmatrix} 1 & 0 & 0 \\ 0.983167069 & 1 & 0 \\ 0.907505346 & 0 & 1 \end{Vmatrix}$$

$$H' = \begin{Vmatrix} 1 & 0.983167069 & 0.907505346 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

$$H^* = \begin{Vmatrix} 1 \\ 0.983167069 \\ 0.907505346 \end{Vmatrix}$$

8. *Compute the Matrix P**

The initial matrix for the Crout computational method of determining P^* is

$$M = \begin{Vmatrix} W & H^* \\ H' & O \end{Vmatrix}$$

Accordingly

$$P^* = \begin{Vmatrix} 0.000256009 \\ 0.000107283 \\ 0.000040962 \end{Vmatrix}$$

9. *Compute the Matrix Q**

Q^* is computed by post-multiplying P^* by A_{11}, thus

$$Q^* = \begin{Vmatrix} 0.000256009 \\ 0.000107283 \\ 0.000040962 \end{Vmatrix} \begin{Vmatrix} 636.0746 \end{Vmatrix} = \begin{Vmatrix} 0.162840550 \\ 0.068240147 \\ 0.026054704 \end{Vmatrix}$$

In this case the first element of Q^* is the submatrix Q_{11}, i.e.:

$$Q_{11} = \begin{Vmatrix} 0.162840550 \end{Vmatrix}$$

10. *Derive Characteristic Equation and Latent Vectors of Q_{11}*

The coefficients of the characteristic equation are determined by generating Bryan's matrix-vector sequence outlined in Section B of this appendix. However, in this case the sequence degenerates into

$$f_1 = (-1 \quad 0.162840550)$$

Thus the characteristic equation is

$$\lambda - 0.162840550 = 0$$

from which it is apparent that the root is

$$\lambda = 0.162840550$$

The matrix L is therefore

$$L = \begin{Vmatrix} 1 \end{Vmatrix}$$

For this problem the numbers are:

	6176.6150	2019.9922	2029.1739	1.0000	10226.7811
	2019.9922	5698.2226	3487.6591	0.983167069	11206.857067
	2029.1739	3487.6591	7404.8265	0.907505346	12922.567005
	1	0.983167069	0.907505346	0	2.890672
	0	1	0	0	1
	0	0	1	0	1
Sum	10226.7811	11207.857067	12923.567005	2.890672415	

The modified auxiliary is:

					Sum and Check
	6176.6150	0.327038710	0.328525236	0.000161901	1.655725847
					0.655725847
	2019.9922	5037.606956702	0.560591708	0.000130246	1.560721954
					0.560721954
	2029.1739	2824.040685378	5155.057874552	0.000040962	1.000040962
					0.000040962
	1	0.656128359	0.211159995	0.000256009	0.000256009
	0	1	-0.560591708	0.000107283	0.000107283
	0	0	1	0.000040962	0.000040962
Sum	10226.7811	7863.303770439	5155.708442537	0.000404253	
Check	10226.7811	7863.303768673	5155.708440562	0.000404253	

and the product B_1L is also

$$B_1L = \| 1 \|$$

since $B_1 = 1$ Thus $S^* = \| 1 \|$ also.

11. *Compute the Matrix Z*

To compute Z, postmultiply Q^* by S^* getting

$$Z = \begin{Vmatrix} 0.162840550 \\ 0.068240147 \\ 0.026054704 \end{Vmatrix} \quad \begin{Vmatrix} 1 \end{Vmatrix}$$

$$= \begin{Vmatrix} 0.162840550 \\ 0.068240147 \\ 0.026054704 \end{Vmatrix}$$

12. *Compute the Matrix V*

To compute V, postmultiply J' by Z getting

$$V = \begin{Vmatrix} 1 & -0.983167069 & -0.907505346 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \cdot \begin{Vmatrix} 0.162840550 \\ 0.068240147 \\ 0.026054704 \end{Vmatrix}$$

$$= \begin{Vmatrix} 0.072104302 \\ 0.068240147 \\ 0.026054704 \end{Vmatrix}$$

13. *Conventionalize V*

The conventionalized coefficients are determined by dividing each element by the largest, i.e. 0.072104302. This gives

$$V = \begin{Vmatrix} 1.000000000 \\ 0.946408815 \\ 0.361347427 \end{Vmatrix}$$

14. *Verify Results*

To verify results compute $V'AV$ and see if it equals $\lambda V'WV$. In this case $V'AV = 3244.223888$ and $\lambda V'WV = 3244.223927$ which agree within the significance of the digits which were carried in computation.

IV. PROPORTIONALITY OF DISCRIMINANT AND REGRESSION FUNCTIONS IN THE TWO-GROUP CASE

In the case of two groups, it is possible to treat the group classification as a criterion variate by assigning arbitrary, but different, values to membership in each group. After this is done, discrimination between groups may be accomplished by determining the linear combination of the x_j ($j = 1, 2, \ldots, n$) variates which has the highest correlation with the criterion. This may be done by any one of several ways of computing the coefficients of a multiple regression equation. Fisher (3, pp. 376–386) has shown that these coefficients are proportional to the coefficients of the discriminant function for this case. Since the proportionality factor does not affect discrimination, discriminant and regression analyses lead to similar classification of individuals in the two-group case. The non-applicability of regression analysis to problems involving more than two groups is obvious.

In order to illustrate the proportionality of the multiple regression function to the discriminant function in the two-group case, the regression on the group variate of the same three variates which were used in the preceding section was computed. Data for calculation of the correlation matrix are given in Subsections C–2 and C–6 of this appendix. The correlations are:

Variate	2	12	15	Group
2	-	0.403372650	0.354636908	-0.308558734
12		- -	0.573178778	-0.312076663
15			- -	-0.257041301
Group				- -

The abbreviated Doolittle method (regular Crout auxiliary matrix followed by back solution) yields the following regression coefficients for predicting group membership:

Variate	Regression Coefficient
2	0.023869111
12	0.022589942
15	0.008625045

In conventional form these coefficients become

1.00000000
0.94640902
0.36134776

which agree with the coefficients in subsection C–13 to the fifth place and demonstrate the proportionality of discriminant and regression analyses in the two-group case.

V. ILLUSTRATION OF DISCRIMINANT ANALYSIS—THREE GROUPS AND FOUR VARIATES

The two-group and three-variate illustration presented in Section C had only one discriminant function, as is to be expected, and resulted in a degeneration of Bryan's procedure for determining the characteristic equation. The simplest illustration having potential multiple discriminators, two to be exact, and requiring Bryan's procedure for calculation of the characteristic equation is the three-group and four-variate case. Since an illustration of this case was given by Bryan in his dissertation, it is reproduced here in place of one which could have been computed for Air Force data.

The Kuder Preference Record was given to a group of Harvard students in the middle of their

freshman year. At the end of the year, the freshman chose a field of concentration without being informed of their scores on the Kuder Preference Record. Sophomore grades in the field of concentration were used in classifying these students as successful concentrators in one of five areas. In his illustration, Bryan chose four Kuder Preference Record scores, computational (X_1), scientific (X_2), literary (X_3), and musical (X_4) and three fields of concentration, social relations (group 1), economics (group 2), and government (group 3).

The computational routine was as follows:

1. *Compute Matrix A*

The first step in deriving the three-group discriminant functions is to calculate and check the elements of A and W. To be sure, the arithmetic can be shortened if the elements of A are arrived at by subtracting W from T, but this procedure sacrifices the check available through an independent determination of each matrix.

The elements of A are computed independently by the procedure described in Subsection C-3 of this appendix. Since this method has been illustrated previously and no change in procedure is required for this illustration, only the A matrix is recorded (1, pp. 171–174).

It is:

$$A = \begin{Vmatrix} 4.906175 & 0.053976 & -1.306168 & -0.657741 \\ 0.053976 & 4.193953 & -2.313869 & 0.264608 \\ -1.306168 & -2.313869 & 1.608709 & 0.026039 \\ -0.657741 & 0.264608 & 0.026039 & 0.105802 \end{Vmatrix}$$

Bryan divided each element in A by 1000 and each element in W by 10000. For simplicity, he still designated the resulting matrices as A and W. As a result of this operation the computed latent roots will be ten times their actual value, but the latent vectors will not be altered.

2. *Compute Matrix W*

Bryan computed matrix W by the procedure described and illustrated in Subsection C-4. Since the addition of a group and a variate does not alter the general scheme of computation only matrix $\hat{W}$ is recorded here (1, pp. 173–174). It is:

$$W = \begin{Vmatrix} 2.511592 & 0.954852 & 0.151910 & -0.436128 \\ 0.954858 & 3.907582 & -1.052430 & -0.666360 \\ 0.151950 & -1.052430 & 5.107310 & 0.293423 \\ -0.436128 & -0.666360 & 0.293423 & 1.738118 \end{Vmatrix}$$

3. *Compute Matrix T as a Check*

The matrix T was computed as described and illustrated in Subsection C-5. Checks were applied by determining:

$$t_{ij} = a_{ij} + w_{ij}$$

for all i and j combinations (1, pp. 173–174).

4. *Compute Matrix B^3*

Symbolically, the initial matrix is

$$M = \left\Vert \begin{array}{cc:cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hdashline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right\Vert$$

Numerically, the initial matrix M and the modified auxiliary matrix K are

$$M = \left\Vert \begin{array}{cc:cc} 4.906175 & 0.053976 & -1.306168 & -0.657741 \\ 0.053976 & 4.193953 & -2.313869 & 0.264608 \\ \hdashline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right\Vert$$

$$K = \left\Vert \begin{array}{cc:cc} 4.906175 & 0.011002 & -0.256229 & -0.134064 \\ 0.053976 & 4.193359 & 0.548367 & 0.064827 \\ \hdashline 1 & -0.011002 & -0.260196 & -0.134777 \\ 0 & 1 & -0.548367 & 0.064827 \end{array} \right\Vert$$

The matrix B is

$$B = \begin{Vmatrix} -.260196 & -.134777 \\ -.548367 & .064827 \end{Vmatrix}$$

As a stringent check, we reconstruct the submatrices of A.

$$A_{11}B = \begin{Vmatrix} -1.306166 & -0.657740 \\ -2.313870 & 0.264607 \end{Vmatrix}$$

$$B'A_{11}B = \begin{Vmatrix} 1.608709 & 0.026040 \\ 0.026040 & 0.105802 \end{Vmatrix}$$

These show excellent agreement with their theoretical equals A_{12} and A_{22}.

5. *Write the Transformation Matrices J, H, H**

$$J = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ .260196 & .548367 & 1 & 0 \\ .134777 & -.064827 & 0 & 1 \end{Vmatrix}$$

$$H = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -.260196 & -.548367 & 1 & 0 \\ -.134777 & .064827 & 0 & 1 \end{Vmatrix}$$

It is to be noted that the transpose of B and not B itself is used in J and H. Since H^* consists of merely the first two columns of H it is unnecessary to write it down separately.

6. *Compute the Matrix P**

Symbolically the initial matrix is

$$M = \left\Vert \begin{array}{c:c} W & H^* \\ \hdashline H' & 0 \end{array} \right\Vert$$

[3]The information in Subsections E-4 through E-12 is practically a verbatim copy from Bryan's dissertation (1, pp. 175–190). Slight editing has been performed here and there to make the style consistent with that of this report.

The numerical matrices are as follows:

*INITIAL MATRIX FOR P**

$$\begin{vmatrix} 2.511592 & 0.954858 & 0.151910 & -0.436128 & 1 & 0 \\ 0.954858 & 3.907582 & -1.052430 & -0.666360 & 0 & 1 \\ 0.151910 & -1.052430 & 5.107810 & 0.293423 & -0.260196 & -0.548367 \\ -0.436128 & -0.666360 & 0.293423 & 1.738118 & -0.134777 & 0.064827 \\ 1 & 0 & -0.260196 & -0.134777 & 0 & 0 \\ 0 & 1 & -0.548367 & 0.064827 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{vmatrix}$$

*AUXILIARY MATRIX FOR P**

$$\begin{vmatrix} 2.511592 & 0.380180 & 0.060484 & -0.173646 & 0.398154 & 0 \\ 0.954858 & 3.544564 & -0.313207 & -0.141?1? & -0.107257 & 0.282122 \\ 0.151910 & -1.119183 & 4.750405 & 0.024318 & -0.092572 & -0.049503 \\ -0.436128 & -0.500553 & 0.163025 & 1.526102 & 0.000172 & 0.134994 \\ 1 & -0.380180 & -0.439785 & 0.000273 & 0.479640 & -0.085451 \\ 0 & 1 & -0.235160 & 0.214114 & -0.085451 & 0.322667 \\ 0 & 0 & 1 & -0.024318 & -0.092578 & -0.054136 \\ 0 & 0 & 0 & 1 & 0.000172 & 0.134994 \end{vmatrix}$$

Accordingly

$$P^* = \begin{vmatrix} 0.479640 & -0.085451 \\ -0.085451 & 0.322667 \\ -0.092578 & -0.054136 \\ 0.000172 & 0.134994 \end{vmatrix}$$

7. Compute the Matrix Q^*

$$Q^* = P^* A_{11} = P^* \begin{vmatrix} 4.906175 & 0.053976 \\ 0.053976 & 4.193953 \end{vmatrix}$$

$$= \begin{vmatrix} 2.348585 & -0.332488 \\ -0.401821 & 1.348638 \\ -0.457126 & -0.232041 \\ 0.008130 & 0.566168 \end{vmatrix}$$

8. Derive the Characteristic Equation and the Latent Roots and Vectors of Q_{11}

$$Q_{11} = \begin{vmatrix} 2.348585 & -0.332488 \\ -0.401821 & 1.348638 \end{vmatrix}$$

$$f_1 = (-1 \quad 2.348585)$$

$$B_2 = \begin{vmatrix} 0 & 0.332488 \\ -1 & 2.348585 \end{vmatrix}$$

$$f_2 = (1 \quad -3.697223 \quad 3.033790)$$

Characteristic Equation:

$$\lambda^2 - 3.697223\lambda + 3.033790 = 0$$

Latent Roots:

$$\lambda_1 = 2.467946\,, \qquad \lambda_2 = 1.229278$$

$$L = \begin{vmatrix} 2.467946 & 1.229278 \\ 1 & 1 \end{vmatrix};$$

$$\Lambda = \begin{vmatrix} 2.467946 & 0 \\ 0 & 1.229278 \end{vmatrix}$$

$$B_2 L = \begin{vmatrix} 0.332488 & 0.332488 \\ -0.119361 & 1.119307 \end{vmatrix}$$

$$S^* = \begin{vmatrix} 1.000000 & 0.297048 \\ -0.358993 & 1.000000 \end{vmatrix}$$

Here S^* represents the conventionalized matrix of latent vectors of Q_{11}.

9. Compute the Matrix Z

$$Z = Q^* S^* = \begin{vmatrix} 2.467946 & 0.365154 \\ -0.885973 & 1.229278 \\ -0.373825 & -0.367829 \\ -0.195120 & 0.568583 \end{vmatrix}$$

$$\text{Check: } S^*\Lambda = \begin{vmatrix} 2.467946 & 0.365155 \\ -0.885975 & 1.229278 \end{vmatrix}$$

The first two rows of Z are in satisfactory agreement with $S^*\Lambda$.

10. Compute and Conventionalize the Matrix V. Coefficients of the Discriminant Functions

$$J'Z = \begin{vmatrix} 2.344381 & 0.346078 \\ -1.078317 & 0.990713 \\ -0.373825 & -0.367829 \\ -0.195120 & 0.568583 \end{vmatrix}$$

$$V = \begin{vmatrix} 1.000000 & 0.349322 \\ -0.459958 & 1.000000 \\ -0.159456 & -0.371277 \\ -0.083229 & 0.573913 \end{vmatrix}$$

The first column of V represents the coefficients of the first discriminant function; the second

column, the coefficients of the second. The matrix V was computed from J′Z by dividing each column by its largest element.

11. *Verify Results*

Before proceeding to the numerical evaluation of the discriminant functions for individual students, the vectors should be tested against their analytical definitions. To this end we compute AV, WV, and WVΛ, comparing the latter with the first. Then we compute V′AV and V′WV to check the vanishing of the off-diagonal terms. Finally we test the latent roots by taking the quotients of corresponding diagonal elements of V′AV and V′WV. It will be seen that to the number of figures carried, all of these tests are satisfactorily met.

$$AV = \left| \begin{matrix} 4.906175 & .053976 & -1.306160 & -.657761 \\ .053976 & 4.193953 & -2.313869 & .264608 \\ -1.306168 & -2.313869 & 1.608709 & .026039 \\ -.657761 & .264608 & .026039 & .105802 \end{matrix} \right| V$$

$$= \left| \begin{matrix} 5.144368 & 1.875275 \\ -1.528129 & 5.223756 \\ -.500571 & -3.352475 \\ -.792407 & .085898 \end{matrix} \right|$$

$$WV = \left| \begin{matrix} 2.511592 & .954858 & .151910 & -.436128 \\ .954858 & 3.907582 & 1.052430 & -.666360 \\ .151910 & -1.052430 & 5.107310 & .293423 \\ -.436128 & -.666360 & .293423 & 1.738115 \end{matrix} \right| V$$

$$= \left| \begin{matrix} 2.084473 & 1.525512 \\ -.619189 & 4.249445 \\ -.202829 & -2.727192 \\ -.322080 & .069875 \end{matrix} \right|$$

$$WV\Lambda = WV \left| \begin{matrix} 2.467946 & 0 \\ 0 & 1.229278 \end{matrix} \right| = \left| \begin{matrix} 5.144367 & 1.875278 \\ -1.538125 & 5.223749 \\ -.500571 & -3.352477 \\ -.792408 & .085897 \end{matrix} \right|$$

$$V'AV = \left| \begin{matrix} 1.000000 & -.459958 & -.169456 & -.083229 \\ .349322 & 1.000000 & -.371277 & .573913 \end{matrix} \right| AV$$

$$= \left| \begin{matrix} 5.998013 & -.000010 \\ -.000011 & 7.172826 \end{matrix} \right|$$

$$V'WV = \left| \begin{matrix} 2.428339 & -.000003 \\ -.000003 & 5.834986 \end{matrix} \right|$$

__Quotient of Corresponding Diagonals__

Elements of $V'\Lambda V$ and $V'WV$

First Quotient = 2.467947 λ_1 = 2.467946
Second Quotient = 1.229279 λ_2 = 1.229278

12. *Computation of Discriminant Function Scores for Each Student*

The discriminant functions, hereafter referred to as V_1 and V_2, were evaluated for each student. The numerical equivalents of $V'\Lambda V$ and $V'WV$ were computed from the individual values of the discriminant functions. The results are shown below.

$$V'\Lambda V = \begin{vmatrix} 5993.00 & -0.01 \\ -0.01 & 7172.69 \end{vmatrix};$$

$$V'WV = \begin{vmatrix} 24283.36 & -0.08 \\ -0.08 & 58349.61 \end{vmatrix}$$

The ratios of the diagonal elements of these two matrices agree very well with the corresponding latent roots, account being taken of the factor of ten as previously explained.

$$\frac{5993.00}{24283.36} = .246795 \qquad \lambda_1 = .246795$$

$$\frac{7172.69}{58349.61} = .122926 \qquad \lambda_2 = .122928$$

13. *Diagram of Individual Discriminant Function Scores*

V_1 and V_2 were depicted as horizontal and vertical axes of rectangular coordinates and the two values for each student were plotted as a point in this plane. These points have been plotted in a different color for each group and are shown in Figure 2.1 (Appendix C); group 1, the social relations concentrators, by green points; group 2, the economics concentrators, by black points; and, group 3, the government concentrators, by red points.

Some clustering of like-colored points and centroids may be observed in Figure 2.1 (Appendix C). This suggests the possibility of providing guidance in field of concentration selection by means of the four Kuder Preference Record scores. However, the amount of overlap in the three bivariate distributions indicates that such guidance would not be very accurate.

APPENDIX B

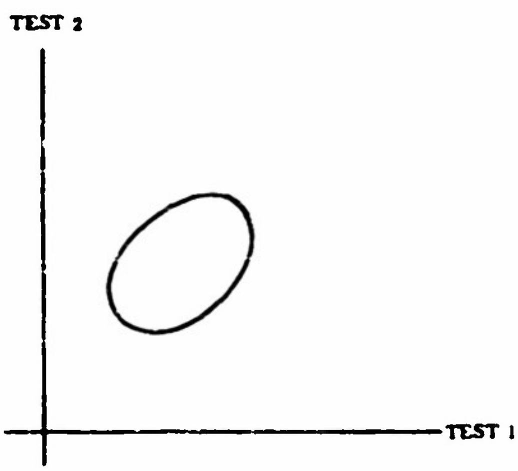
USE OF MULTIPLE DISCRIMINANT FUNCTIONS IN THE OBJECTIVE MEASUREMENT OF MOTIVATION AND TEMPERAMENT[1]

Dr. Phillip J. Rulon

Because the time is rather short, I shall say only three things, and I shan't elaborate on any of them unless there are questions some of you want to ask about some of their details.

First, I want to discuss a little elementary geometry. Let us start with an ordinary scatter diagram as employed in zero-order correlation. To simplify the diagram we will not present the separate entries or tallies in the scatter, but only one of the iso-frequency contours (see Fig. 1).

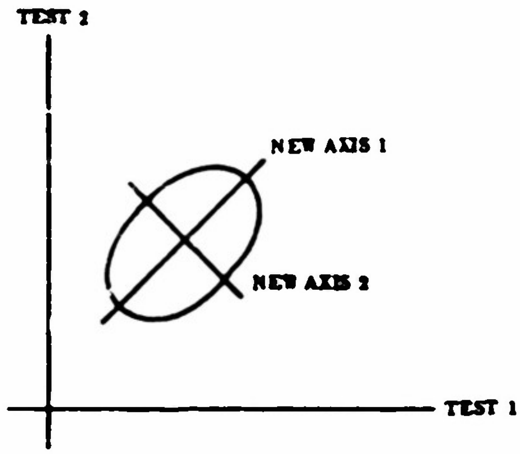
FIGURE 1



This picture shows, then, the scores earned on two tests by a group of individuals. The plane of the diagram is called the test space and since there are two tests, the space is two-dimensional.

Several things can be done with such a portrayal. One of these is the translation and rotation of axes so that the origin of the coordinate system is at the centroid of the scatter and the directions of the coordinate axes correspond to the slopes of the major and minor axes of the iso-frequency ellipse. Figure 2 shows the result of such a translation and rotation.

FIGURE 2



This sort of rearrangement of the portrayal has been employed systematically in factor analysis, and I want to talk about something else. If this scatter shows the scores of a group of doctors, I will add the results of giving the same two tests to a group of lawyers. Then we have two scatters, each represented by an iso-frequency contour as in Figure 3.

In Figure 3 I have sketched in the elliptical axes to show roughly the same factor pattern among the lawyers as among the doctors. In this situation it is transparently obvious that the factor pattern does not distinguish *between* the doctors and the lawyers. It is almost as clear that no amount of rotation, translation, or other transformation procedure applied first to the doctors and then to the lawyers can ever make comparisons *between* doctors and lawyers. If someone asks whether it is Test 1 or Test 2 which best distinguishes *between* doctors *and* lawyers, we will never be able to answer from

[1]This report is part of Human Resources Research Center *Conference Report* 51-3 (8, pp. 82-95).

FIGURE 3

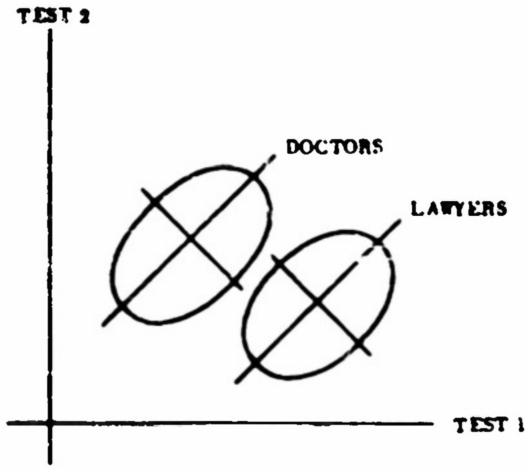
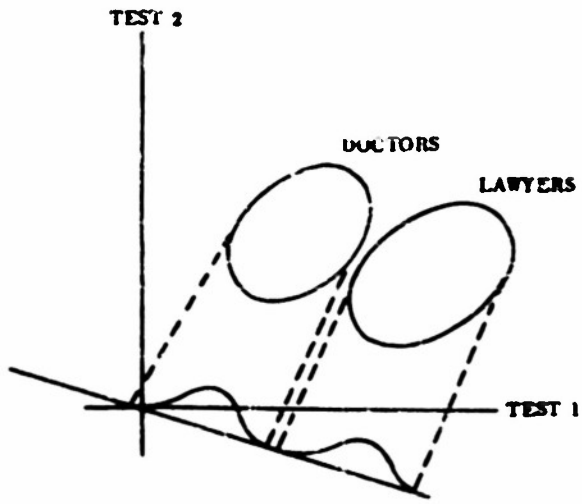


FIGURE 4



any correlation study of the doctors' scores, or any correlation study of the lawyers' scores, or any comparison between such correlation studies.

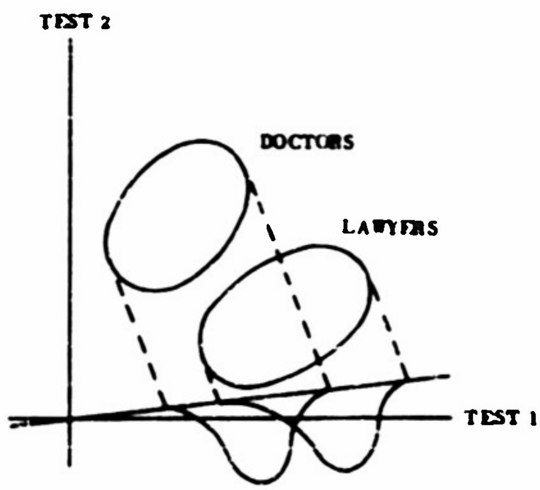
What I have said refers to two tests, but it is just as true if we have given 5 or 12 or any other number of tests. No factor analysis of doctors' scores from 17 tests will tell which of those 17 are distinguishing between doctors and lawyers or between doctors and any other group whose scores were not entered in the analysis. To make this clear beyond peradventure of a doubt, let's look back at Figure 2 where the doctors' scores have been factor analyzed. In this figure there is nowhere any information about lawyers, and yet all the information necessary for factor analysis of doctors' scores is there.

When we come to the situation shown in Figure 3 where we *do* have data on both doctors and lawyers, we ignore the difference between Figure 2 and Figure 3 if we propose correlation studies in the separate groups as a way of studying the differences between the groups.

The way to study the differences between the groups in terms of the tests given them is to use the Fisher discriminant function. In the case of two groups there is one degree of freedom between groups, and the discrimination can be shown in one dimension—that is, on a straight line. Let's draw a line on Figure 3 and study the projections of the scores on that line (see Figure 4).

On this added sloping line I have sketched a pair of normal univariate distributions to show the distributions (along the line) of the projections of the groups onto the line. Clearly the shapes of these distributions depend upon the slope of the line. This can be seen by comparing Figure 4 with Figure 5, in which the latter distributions on the line overlap each other and show a lot of scatter in each distribution.

FIGURE 5



Apparently there will be one slope for the line where the two groups are most clearly separated. Figure 4 shows a more nearly optimum slope of the line than Figure 5 does. The optimal slope is where the separation of the groups along the line is a maximum compared to the variance within the groups, the latter variance also being

FIGURE 6

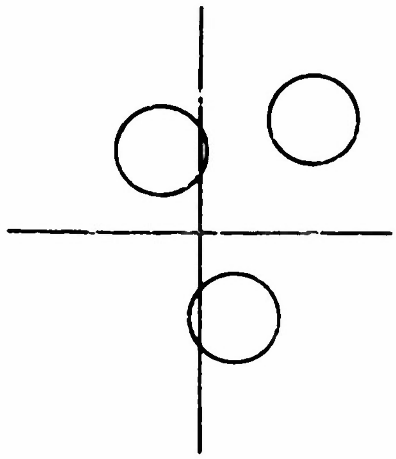
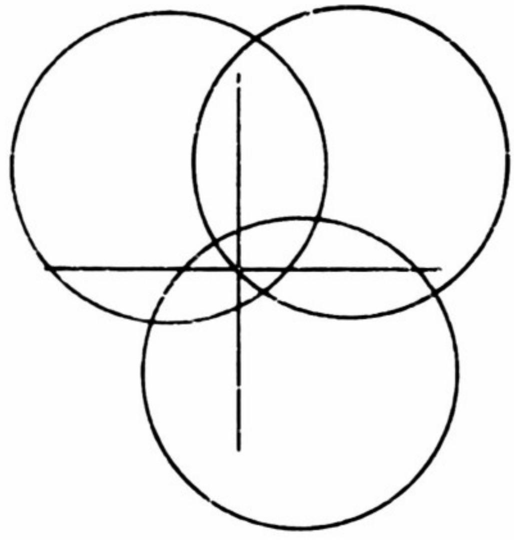


FIGURE 7



measured along the line. The Fisher discriminant function gives the projections when the line is optimally oriented.

We have discussed the case for two tests. But if the doctors and lawyers were given 5 or 12 or any other number of tests, the problem is the same, and so is the solution. If 17 tests are given, the scores form two swarms in 17–space, and we have to orient a line in that space in such a way that the projections onto it form two univariate distributions with maximum separation compared to the internal variances.

All this has concerned two groups, as for example, doctors and lawyers. If a group of engineers is added, there are two degrees of freedom between groups, and to represent these the groups must be projected onto a plane instead of onto a line. If 17 tests are given to the groups, the problem is to orient a plane in the 17–space in such a way as to maximize the separation of the bivariate distributions of the projections on that plane. The separation is to be maximized as compared to the within groups' variance, of course. Figures 6 and 7 show two possible outcomes of such an analysis.

These figures are not to be confused with Figures 3-5 which showed distributions in the test space. Figures 6 and 7 show distributions in the discriminant space, which for three groups is a plane oriented in the (say) 17–space for 17 tests. The plane of Figure 6 is analogous to the sloped line in Figures 4 and 5. So is the plane on which Figure 7 is drawn. Figure 6 shows the outcome when the tests discriminate pretty well between the groups; Figure 7 shows what comes out when the tests do not discriminate the groups so well.

If there are four groups, the discriminant space has three dimensions, and in general if there are G groups the discriminant space will be of G-1 dimensions. So, for any number of groups taking any number of tests, the problem is the same except for the dimensionalities of the relevant spaces.

The details of this extension to G groups taking T tests are not particularly appropriate for elaboration at this time, and I will go on to the second thing I wanted to say, which is something about the implications of all this for the problems raised at this conference. Here I am on something of a spot. If one is asked to say what multiple correlation is good for, he may properly reply that it is useful in all sorts of ways. It is the same with the multiple discriminant analysis. It is good for lots and lots of things. In general, it is the technique of choice whenever we are concerned with relationships between qualitative and quantitative variables.

In the case of the doctors and lawyers the vocation of the individual is a qualitative variable. A man's "score" on profession is either "Doctor" or "Lawyer" and this is a non-quantitative matter. On the other hand the test score earned by the same individual is a quantitative score. Even if it's "pass-fail" it's still quantitative in essence, inasmuch as "pass" is better than "fail," while "Lawyer" is not

FIGURE 8

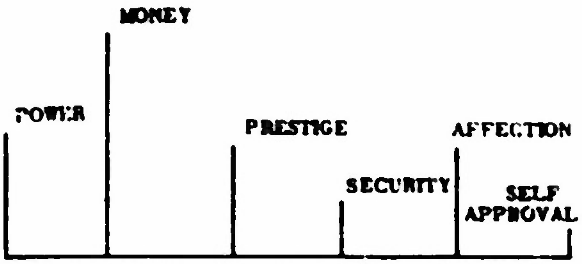
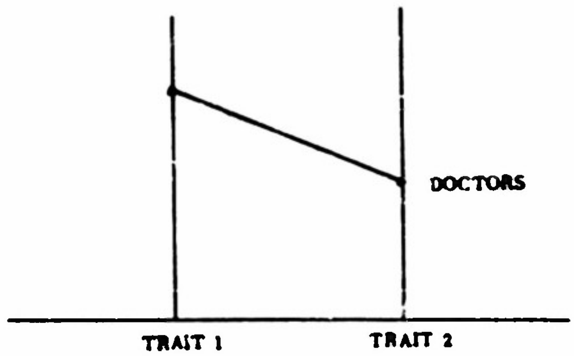


FIGURE 9



regarded as a better score than "Doctor" or "Engineer," as a vocation score.

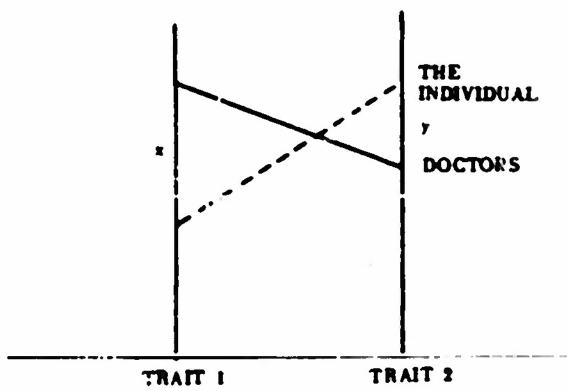
During this conference a number of speakers have discussed problems which involve non-quantitative variables. The occupations or specialties in which the importance of motivation has been emphasized are examples of non-quantitative variables. This is fairly clear. But another and equally central variable is the type of motivation or temperament. Probably all the temperament types, like all the job specialties, can be regarded as qualitative categories and treated with the statistical techniques appropriate to such variables. In the case of motivation, we may appear to have quantitative variables, inasmuch as a worker may be more motivated or less motivated—a quantitative matter.

However, even in this case, the actual human behavior may be forecast and controlled by regarding motivation as a qualitative variable. What I am thinking about is the possibility of a "dominating drive." A man may work eagerly for any of a number of reasons. One man may be motivated by desire for money, another by prestige, another by desire for power, another by need for affection or appreciation, and another by desire for security. The extent to which each of these is effective in any individual may yield a motivation profile, in which the height of the point on each stalk is proportional to the effectiveness of this drive in determining the behavior of the individual. Thus an individual may look like the profile in Figure 8, in which apparently the desire for money is predominant in effectiveness. It is possible that such persons form a distinguishable class whose members are dominated by the money motive. In that case the dominance score for these people would be "money" rather than "security" or "affection." Thus the motivation score becomes qualitative, and the statistics for unordered classes become appropriate in studying the traits and behavior of these people.

I have a suggestion or two for research projects related to these questions but I intended mentioning them under the third section of this presentation. Before leaving this second section I'd like to point out an application of another sort. Dr. Kelley has proposed a measure of how well an individual qualifies for membership in a group. He calls his measure "misfit variance" and so it measures misfit, rather than fit; but, of course, that is no difficulty. Dr. Kelley plots a group on a profile chart, which in a simple case could have just two stalks like the one in Figure 9, for doctors.

Now, an individual is measured on the two traits and his profile is plotted on the same chart, as in Figure 10.

FIGURE 10

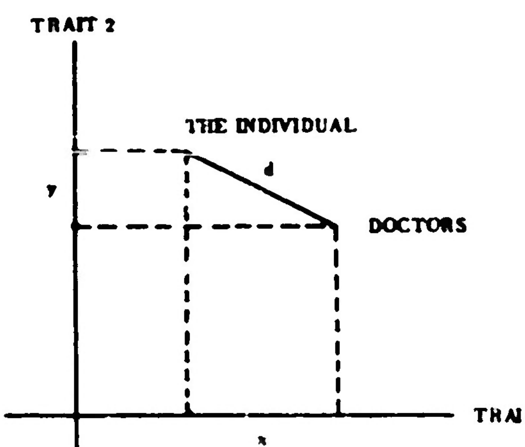


If now we denote the discrepancies by x and y, as in the figure, we can compute the quantity

$x^2 + y^2$; and if the value is large, the individual doesn't fit the group. If the individual profile exactly fitted the group profile, then $d^2 = x^2 + y^2$ would be zero. First let us remark that this is really a sum of squares, and not a variance in the Pearsonian sense, because first the discrepancies do not total zero, and, second, there is no division by N.

I will try to show that under certain conditions this misfit sum of squares is a very rational measure. Let me show this profile in a more logical form. In Figure 11, I have transcribed as well as I could the distances shown in Figure 10. And it is clear in Figure 11 that $d^2 = x^2 + y^2$, by the Pythagorean Proposition. Also it is clear that d is the distance from the individual to the doctors' centroid. This is true only when the traits are plotted orthogonally. And it is meaningful only when meaning can be attached to x and to y. Dr. Kelley has covered both these provisions in his proposed use of the concept, and I need only caution others to do likewise. Otherwise, the misfit sum of squares computed from an irrationally set-out profile chart may yield quite irrational and even misleading results.

FIGURE 11



In leaving this for other matters, I will add only that the same sum of squares has the same meaning when more stalks appear on the profile. We have then merely more dimensions in the Cartesian space and the square of the slant distance is still given by the sum of squares of discrepancies. And this distance has meaning only in terms of the meanings of the discrepancies.

The multiple discriminant analysis derives a set of dimensions (or profile stalks) on which two or more groups may be most clearly differentiated in terms of their profiles, or equivalently in terms of their locations in the discriminant space.

The third thing I want to talk about is specific research on motivation and temperament, employing the type of analysis I have sketched. I am convinced that it is possible to employ temperament characteristics in the management and control of workers, and make use of the various motivating influences upon human behavior, if it is only possible to find out what sort of temperament a worker has or what motivates his behavior. Also it is possible to find out these things, but only by very time-consuming and expensive techniques, such as by individual case study, interview techniques, follow-up studies, and clinical methods.

On the other hand, once these types are identified in terms of the individual's temperaments and the workers are sorted into temperament groups, we can validate a large number of much less expensive group tests, biographical inventories, preference blanks, and the like, using the multiple discriminant technique here in much the same way the multiple correlation techniques were employed in the stanine business.

I should like to urge such a project. If we know that a plumber's eagerness is ascribable to his desire for appreciation of a job well done, then we know how to maintain his eagerness. If an electrician is motivated by desire for money, then we can tell how to motivate him. The plumber and the electrician may have equal criterion scores, one in plumbing and the other in electricity. But they may have quite different motivations and temperaments, and it is these different motivations that we must use to maintain enthusiasm and proficiency on the job. And, of course, two carpenters (or two plumbers, or two electricians) may both be proficient and dependable but for quite different reasons of motivation and temperament. We ought to know this.

A project should be undertaken, in which temperament groups are identified by whatever expensive and time-consuming methods are necessary for this sorting, and then a battery of much less expensive measures validated to accomplish this same classification job in a practical way.

Then we would know what makes Sammy run, and how to make him run if we want him to.

Air Force - Scott AFB. Ill/53-300

Figure 5.1

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group A–Radio Operators (N=334)

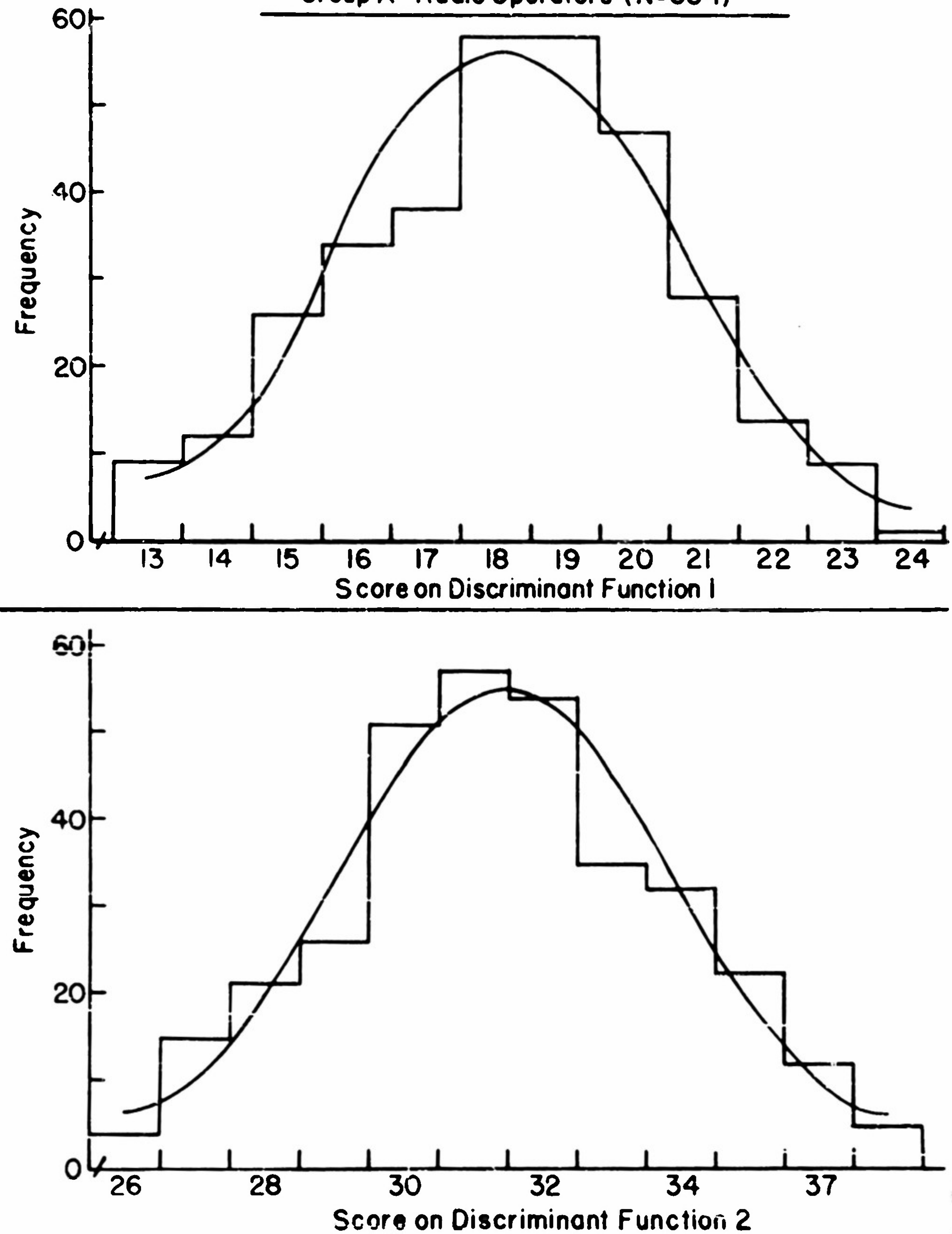


Figure 5.2

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group B — Clerk-Typists (N=1966)

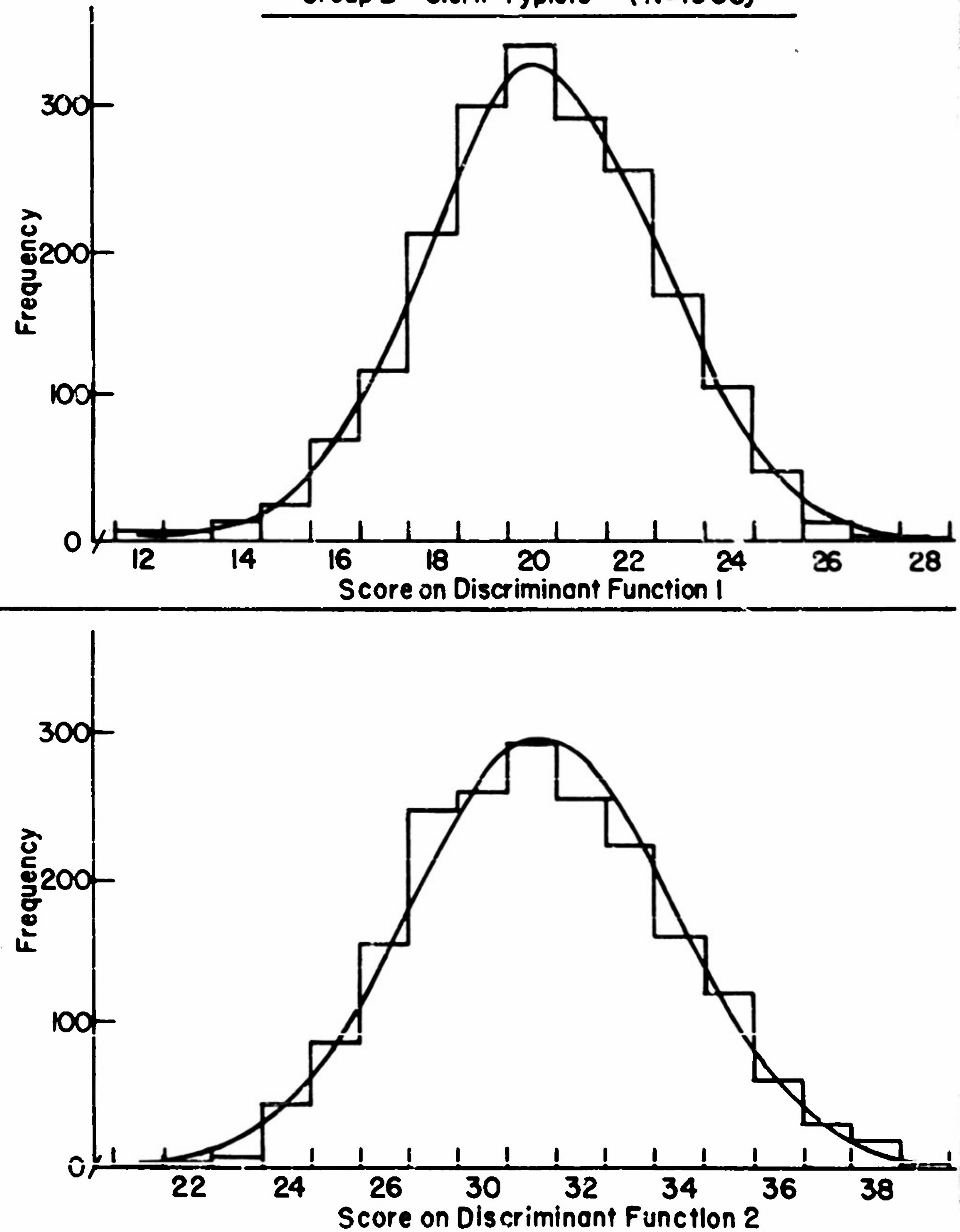


Figure 5.3

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group C-Control Tower Operators (N=620)

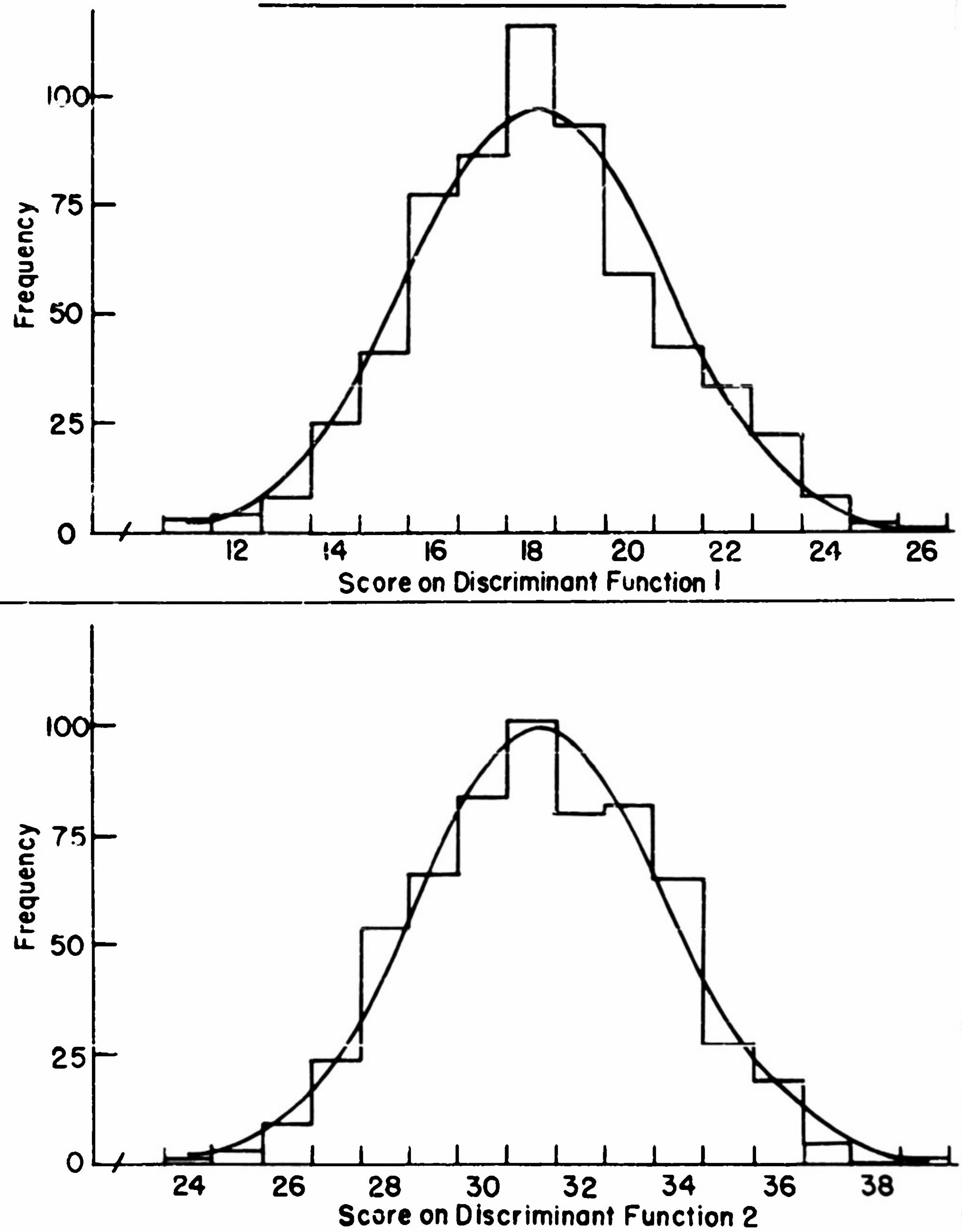
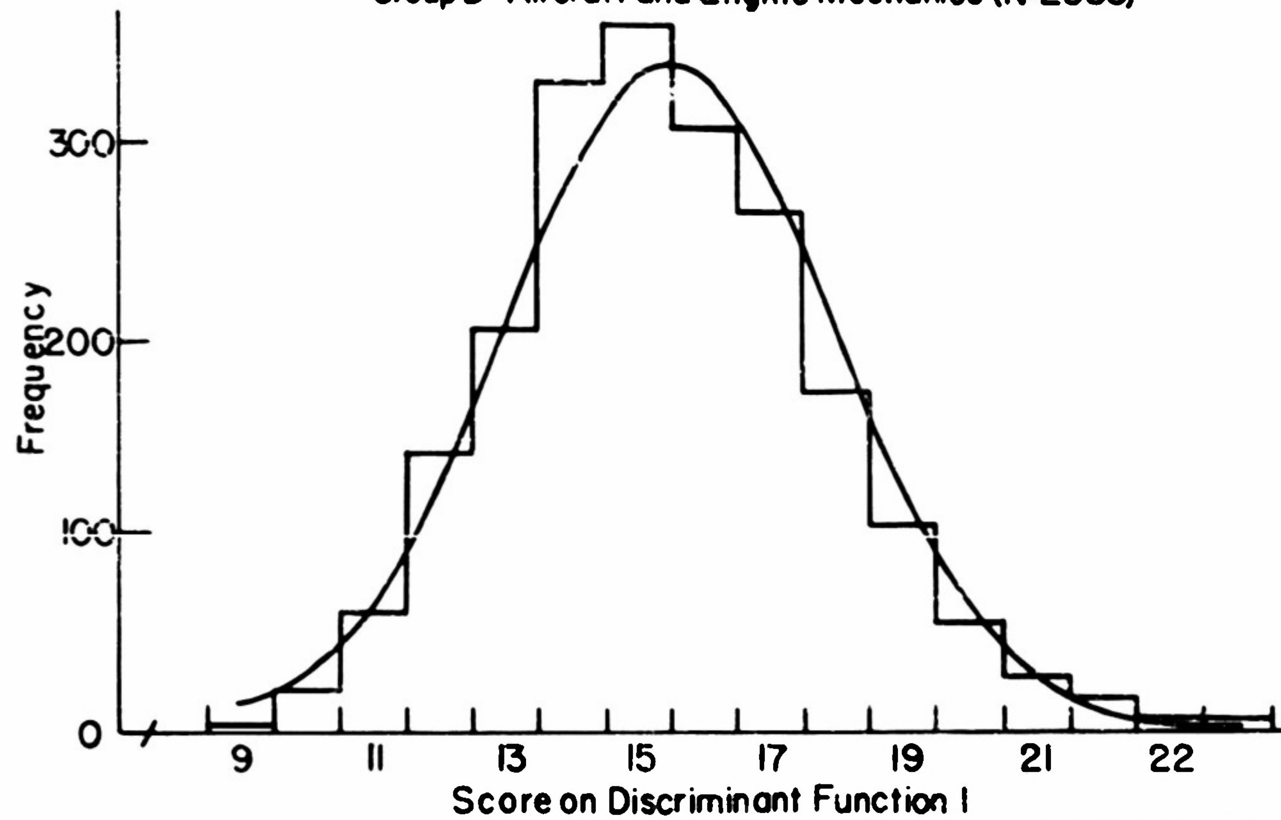


Figure 5.4

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group D-Aircraft and Engine Mechanics (N=2083)



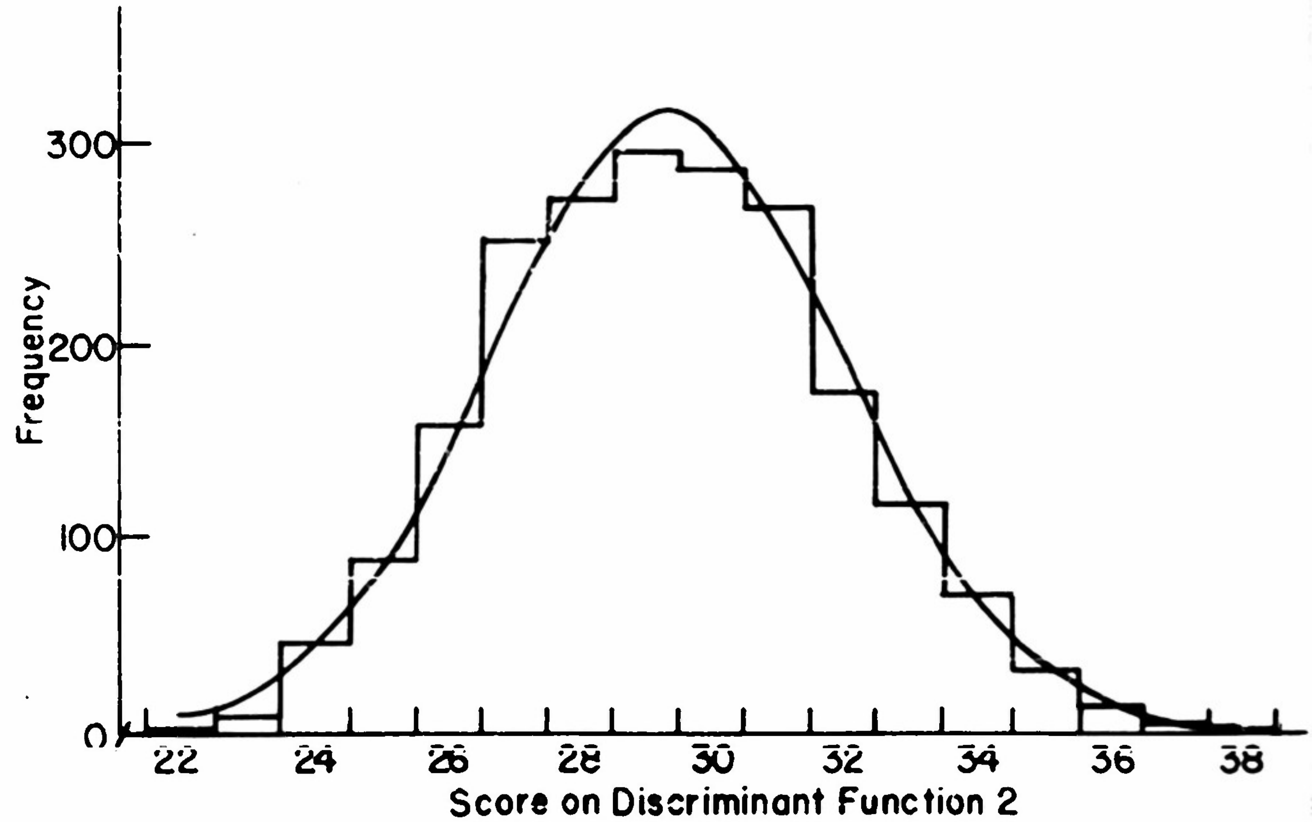


Figure 5.5

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group E – Radio Mechanics (N=514)

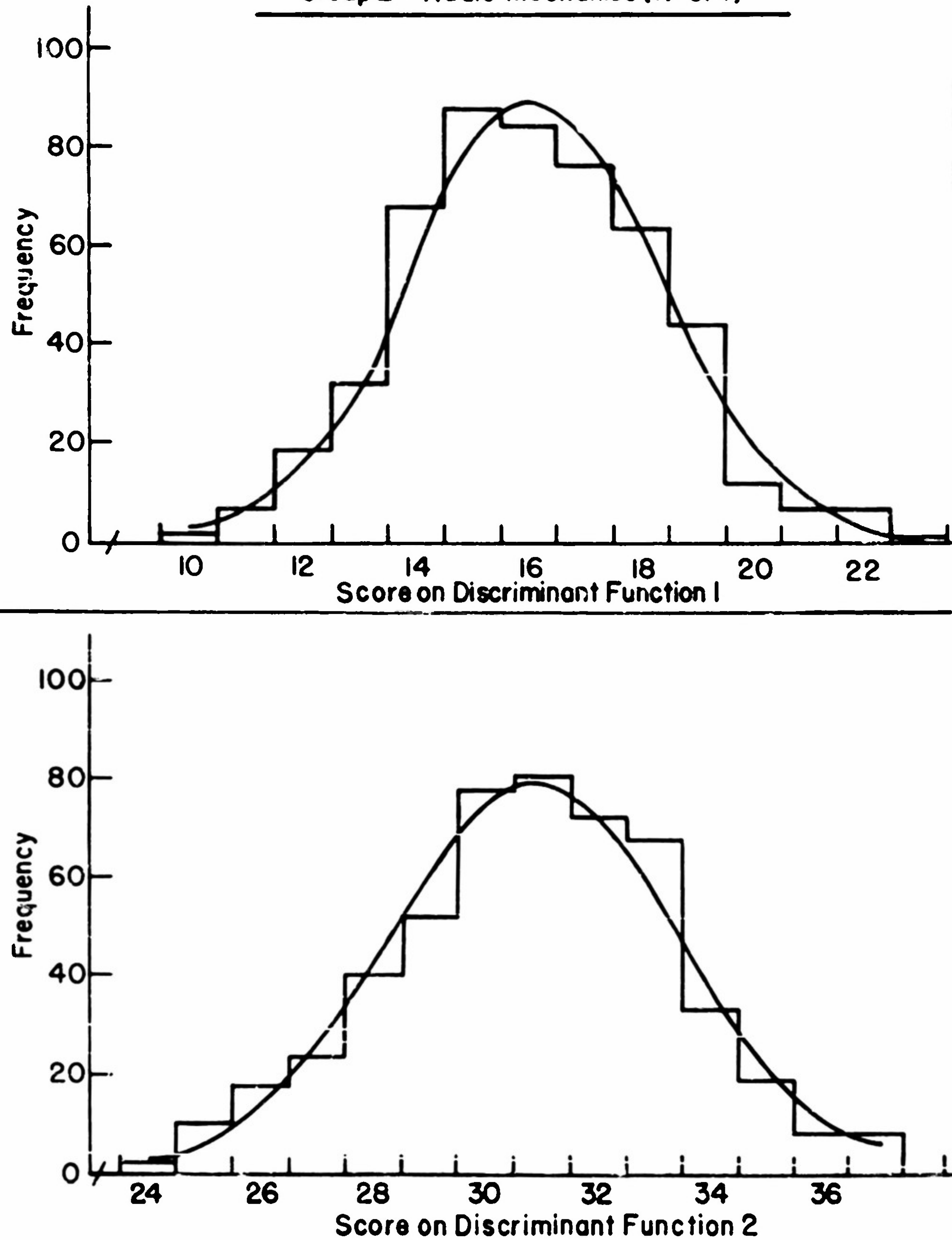


Figure 5.6

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group F - Weather Observers (N=266)

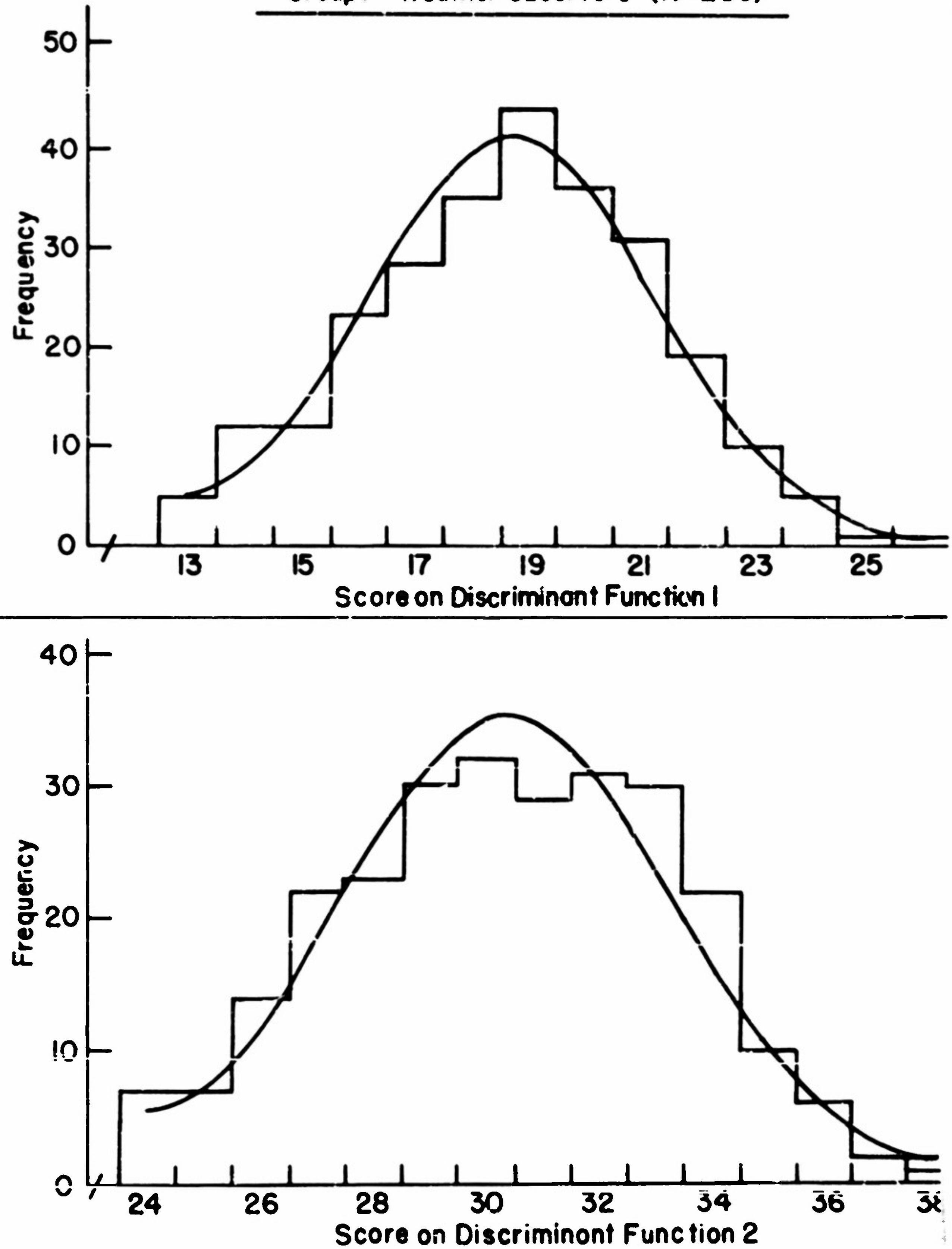


Figure 5.7

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group G – Airplane Sheet-Metal Workers (N=233)

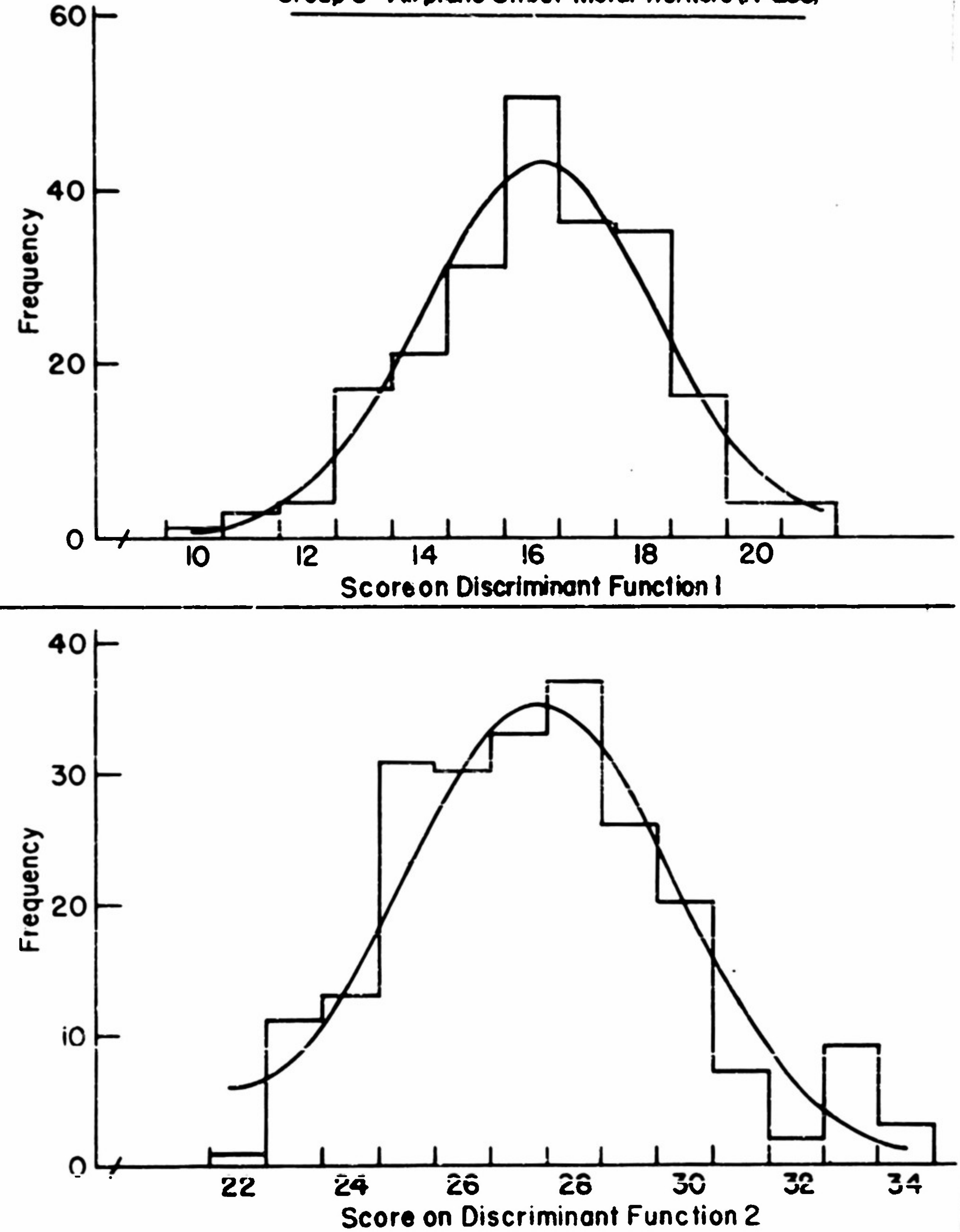


Figure 5.8

Comparison of Discriminant Score Distributions and Normal Distributions of Same Mean and Standard Deviation

Group H— Radar Mechanics (N=99)

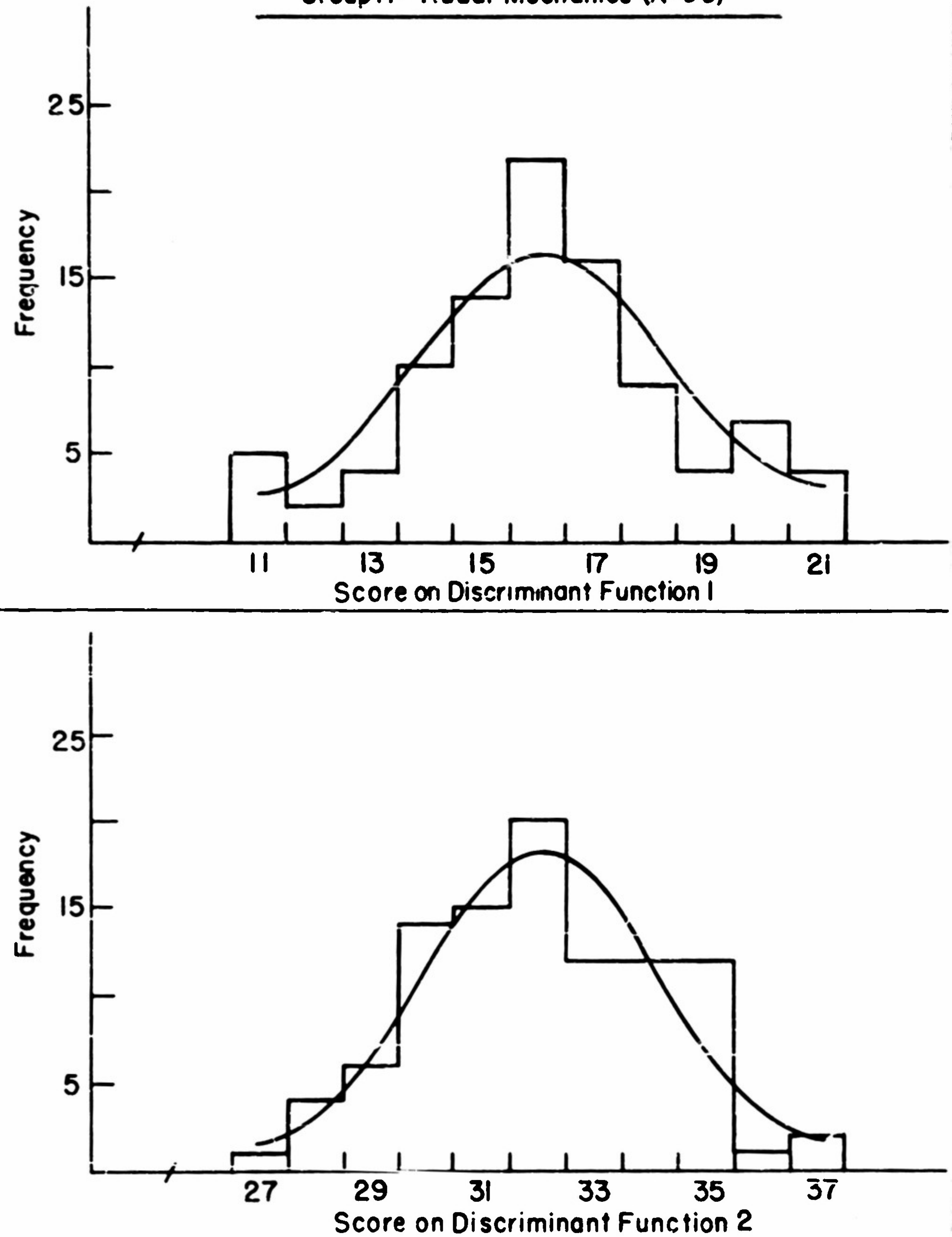


Figure 5.9

Gaussian Contour Ellipses of each Group

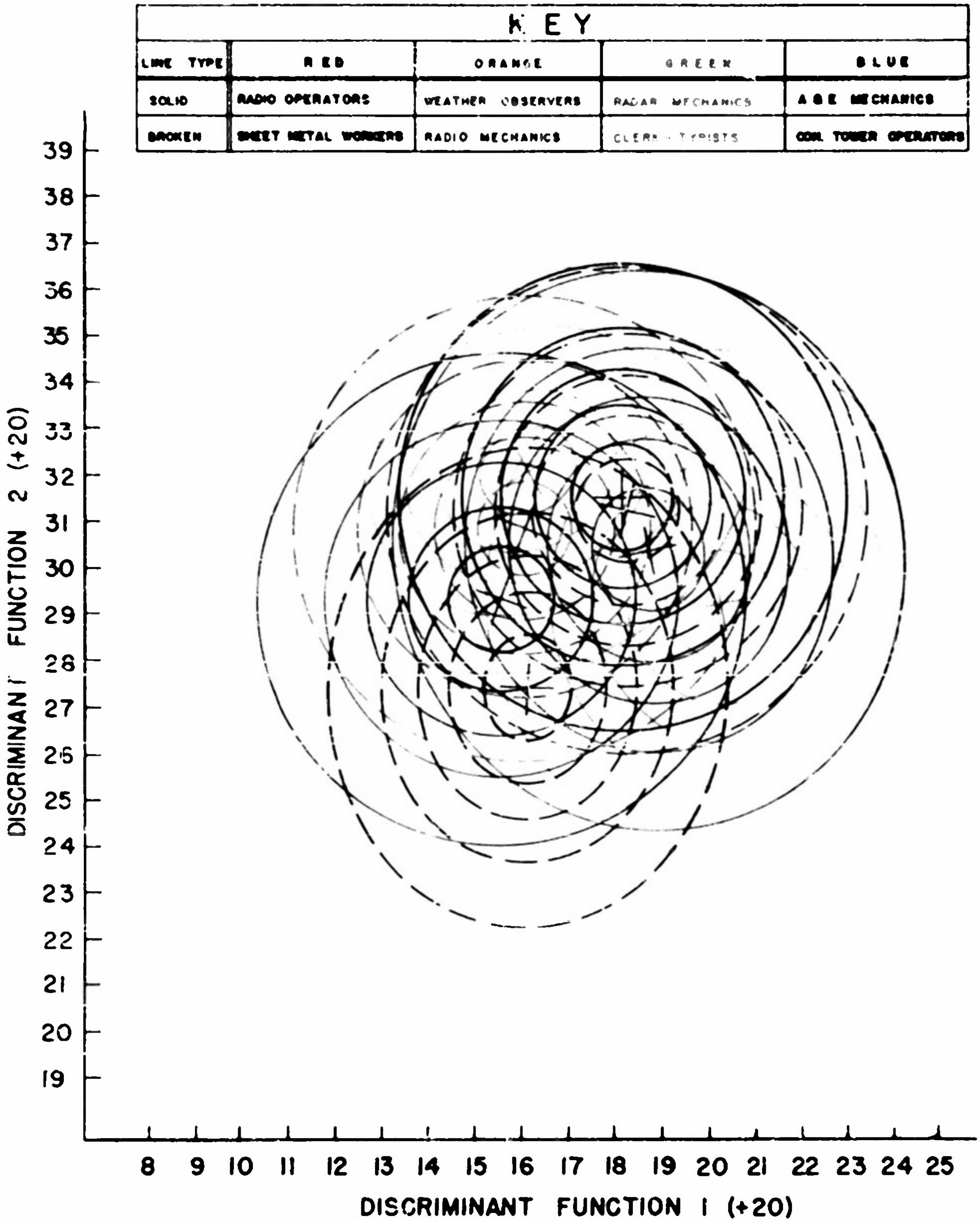


Figure 6.1

Regions Where Centour Scores of Each Group Are Maximal

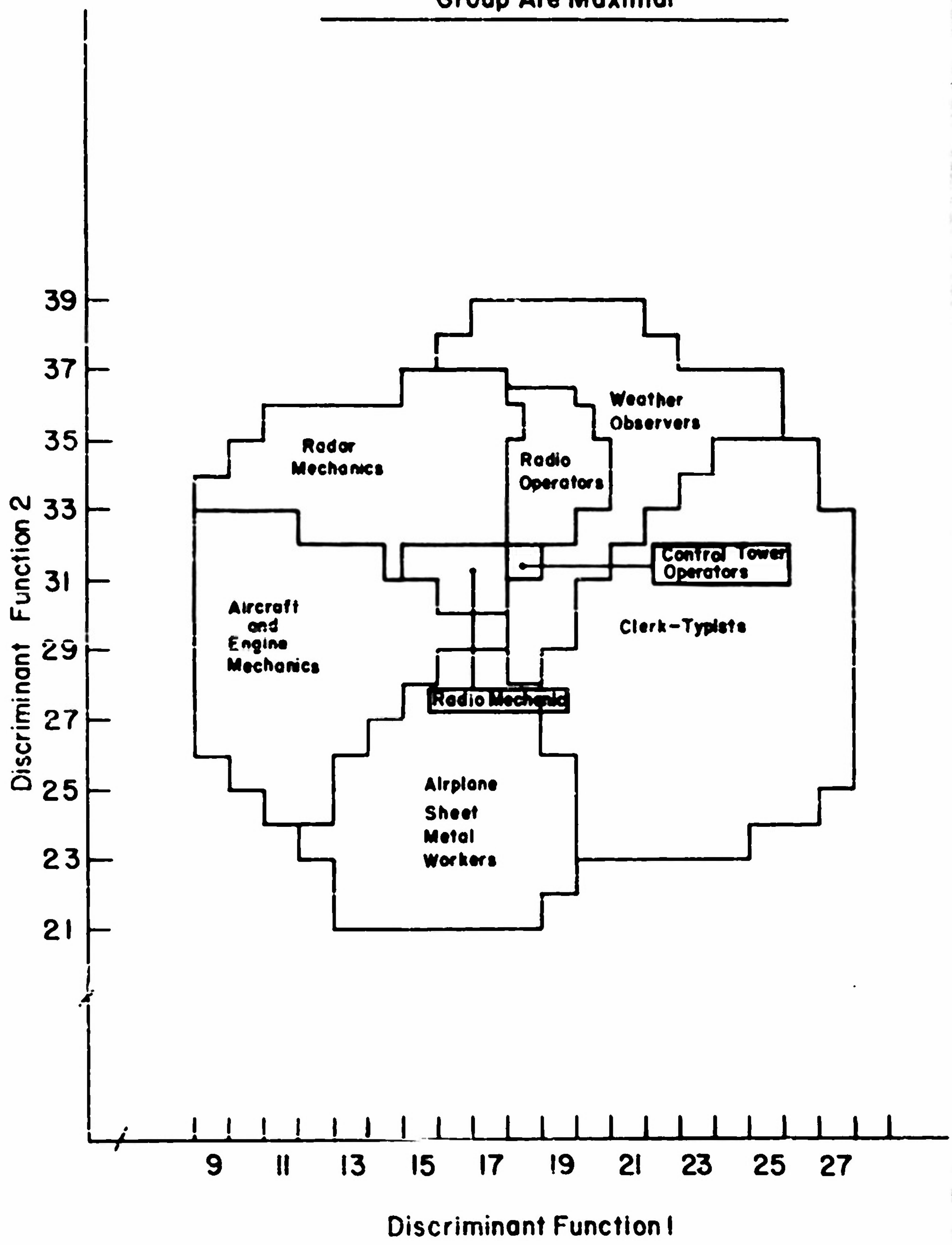


Figure 7.1

Centroids of Discriminant Distributions

for All Groups

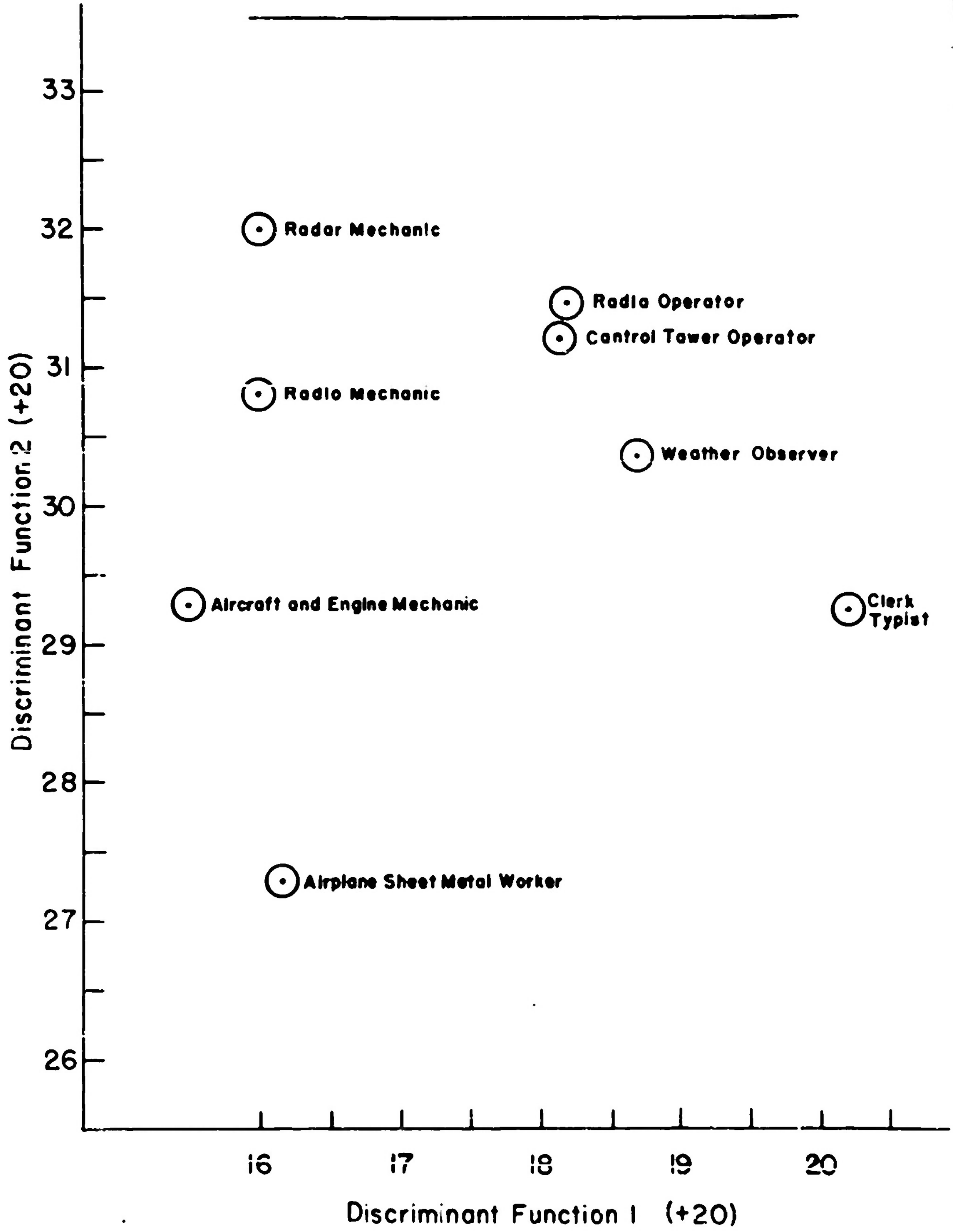
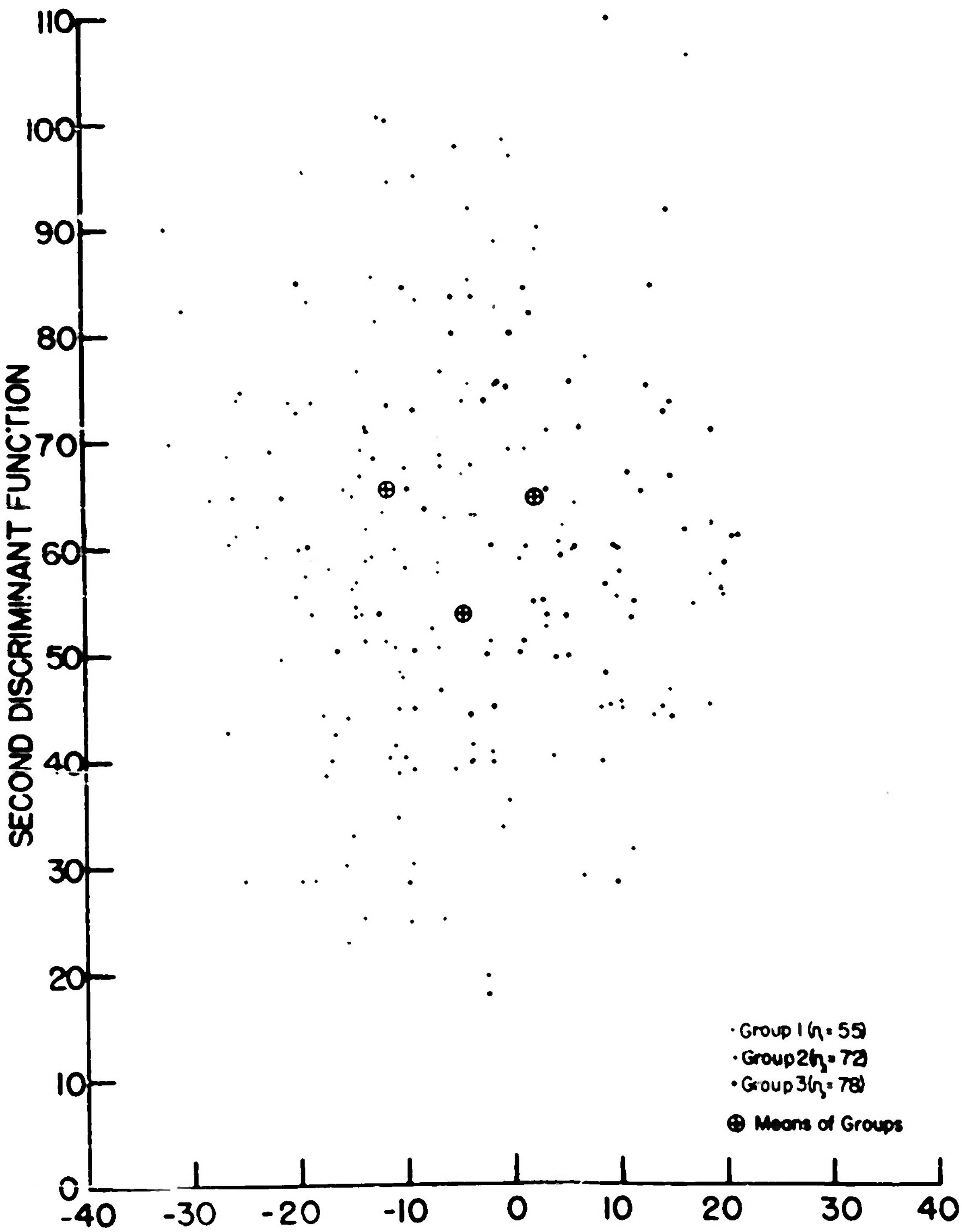


Figure 2.1

THE INDEPENDENT LINEAR COMBINATIONS (DISCRIMINANT FUNCTIONS) OF FOUR OBSERVATIONS WHICH PROVIDE MAXIMUM DISCRIMINATION AMONG THREE GROUPS



(PLOTTED FROM THE WORK OF JOSEPH G. BRYAN)